Estimation of the distribution of change-points with application to fMRI data *

John A D Aston† and Claudia Kirch‡

May 11, 2011

Abstract

Change-point detection in sequences of functional data is examined where the functional observations are dependent and where the distributions of change-points from multiple subjects is required. Of particular interest is the case where the change-point is an epidemic change (a change occurs and then the observations return to baseline at a later time). The case where the covariance can be decomposed as a tensor product is considered with particular attention to the power analysis for detection. This is of interest in the application to functional magnetic resonance imaging (fMRI), where the estimation of a full covariance structure for the three-dimensional image is not computationally feasible. Using the developed methods, a large study of resting state fMRI data is conducted to determine whether the subjects undertaking the resting scan have non-stationarities present in their time courses. It is found that a sizeable proportion of the subjects studied contain a epidemic change. The change-point distribution for those subjects is empirically determined as well as its theoretical properties examined.

Keywords: At most one change; Epidemic change; Functional time series; multidimensional functional data; Resting state fMRI;

AMS Subject Classification 2010: 62M10; 62M40; 62H35

1 Introduction

An increasing number of applications from biology to image sequences in medical imaging involve data that can well be represented as functional time series. This has also lead to a rapid progression of theory associated with functional data, particularly

*This work as well as the position of the second author was financed by the Stifterverband für die Deutsche Wissenschaft by funds of the Claussen-Simon-trust. The first author was also supported by the Engineering and Physical Sciences Research Council (UK) through the CRiSM programme grant and by the project grant EP/H016856/1, and thanks SAMSI for hosting the author during which some of the work was carried out.

†CRiSM, Department of Statistics, University of Warwick, Coventry, CV4 7AL ,UK; J.A.D.Aston@warwick.ac.uk

‡Karlsruhe Institute of Technology (KIT), Institute for Stochastics, Kaiserstr. 89, D – 76133 Karlsruhe, Germany; claudia.kirch@kit.edu

regarding complex correlation structures present within and across many functional observed data. These require methods that can deal both with internal and external dependencies between the observations. Nonparametric techniques for the analysis of functional data are becoming well established (see Ferraty and Vieu [14] or Horváth and Kokoszka [20] for a good overview), and this paper sets out a nonparametric framework for change-point analysis within and across dependent functional data. Two types of change-point alternatives are considered, at-most-one-change (AMOC) and epidemic changes, where the observations having changed return to their original state after some unknown time.

In many applications, sets of functional observations are recorded from a number of subjects, and the distribution of the change-points over all subjects is an item of interest. In addition to giving consistent estimators for the change-points within one set of dependent observations, in Section 4.1 those estimators are used to find the distribution as well as density of the change-points in hierarchical models, where several independent sets of time series including a random change are observed. In this case empirical distribution functions and kernel density estimators based on the estimated change-points for each individual time series yield consistent results (cf. Theorems 4.1 and 4.2).

Tests and estimators are usually based on dimension-reduction techniques, where it is important that the change is not orthogonal to the projection subspace (for details see Section 2.1). Most methodology, including those references given above, chooses this subspace based on principle component analysis assuming a general covariance structure within the functional data. Aston and Kirch [1] showed that estimators and tests can be derived for AMOC and epidemic changes in this case extending work of Berkes *et al.* [4], Aue *et al.* [2] as well as Hörmann and Kokoszka [19]. In the present paper we focus on additional aspects not considered so far, such as estimation of the distribution of change-points in hierarchical models (Section 4.1) as well as practical aspects arising from the consideration of fMRI data. For example, we consider the case of multidimensional separable functional covariance structures in Section 2.3.2. If the underlying covariance structure is indeed separable this leads to a valid estimation of the principle components, but even in the misspecified case it leads to a valid basis selection procedure but with a somewhat different interpretation (cf. Theorem 2.1). This approach is applicable even in situations where the general covariance structure is computationally infeasible, or difficult to estimate due to its very high dimensionality, a situation that arises in the application of brain imaging considered in this work.

The choice of estimator for the covariance is critical for the power analysis in detecting the change. In particular, a large enough separable change will switch the estimated system in such a way that the change is no longer orthogonal to the projection subspace making it detectable (cf. Corollary 2.1), a very appealing feature of the use of functional principal components.

Given its generality, applications for the methodology are fairly widespread. In particular, change-point detection procedures can be useful in image processing applications, where images are taken over time. Of interest in this paper is the use of the derived techniques to find change-points in experiments from Neuroimaging. Recently, change-point analysis has been highlighted as a useful technique in psychological experiments performed with functional Magnetic Resonance Imaging (fMRI) (Lindquist *et al.* [30], Robinson *et al.* [35]) where different subjects react differently to stimuli such as stress or anxiety (as the time of brain state change is much less clearly linked to the stimuli than in an experiment involving movement, for example, where the observed movement and brain activity will be intrinsically linked). A particular type of scan that has recently become very popular is the resting state scan, where subjects are scanned while lying in the scanner "at rest". This data is used to infer connections in the brain which are not due to external stimuli, see for example

Damoiseaux *et al.* [9]. A profound question of interest in analyzing these studies is whether the time series from the experiments are truly stationary or whether they contain level shifts, including segments which return to the original state after some unspecified duration, this activation-baseline pattern being a standard assumption in most fMRI experiments. It is well known that there is dependence between temporal observations within fMRI partly due to scanner effects. Furthermore, multiple subjects are usually scanned, indicating a hierarchical nature of the change-points within the experiments. Current methodology is applied pointwise across spatial locations to find epidemic changes (Robinson *et al.* [35]), requiring a mass univariate approach for this very high dimensional multivariate or functional data, with all the problems that then ensue (particularly of spatially correlated multiple comparisons). By considering each complete image (approximately 10^5 observations) as a single functional observation, we derive a true functional change detection procedure. However, to achieve this computationally, it is necessary to incorporate the three dimensional spatial structure of the observations to estimate the covariance functions required. This motivates our investigation of the multidimensional separable structures derived in this paper.

The paper proceeds as follows. In Section 2 change-point detection procedures for a single functional time series are developed. In particular, in Subsection 2.1, methods for the detection and estimation of change-points for dependent functional observations are reviewed. These methods are presented using an arbitrary orthonormal projection subspace which allows the same general theory to apply regardless of the subspace projection choice. Possible ways of choosing the projection, including issues associated with estimating these projections from the data are detailed in Subsection 2.3 with particular attention to the separable covariance case. In Section 3 practical aspects of small sample testing arising from the application is addressed. The problem of estimating the covariance structure of the subspace has a significant influence on the behaviour of test and estimation procedure, therefore this problem is discussed in detail in Subsection 3.1. Section 3.2 provides a bootstrap implementation for the procedures outlined previously, allowing characterisation of small sample properties for the methodology, as well as overcoming a practical problem of large temporal covariance estimation required for the asymptotic procedure. While the methodology developed thus far deals with change-point procedures for a single functional time series, Section 4 provides estimation procedures for change-point distributions in hierarchical models, i.e. in situations where several independent time series (several subjects) including a change have been observed, accounting for multiple comparisons when testing multiple subjects to determine the overall change distribution across subjects. In Section 5 conclusions are given before the final section gives the details of the proofs.

The data analysis of nearly 200 resting state scans is given throughout the paper as the theory is developed. In Section 2.2 details about the data set are given and examples of projected scans shown indicating that epidemic changes are indeed a good first approximation to the deviation to stationarity that can be expected. Even though the scans are not sparsely represented in terms of basis functions, only a very small number of basis functions are needed to detect change-points in practice (which confirms our theoretic results). In Section 3.3 the test results for the data are reported indicating that $40 - 50\%$ of the resting scans exhibit deviations from stationarity, even after correction for multiple comparisons across subjects. This indicates that substantial care should be taken when combining resting state scans, as non-stationarities will likely be present and these could greatly confound analyses based on correlations for example. Finally, in Section 4.2 the estimators for the position and duration of the change are given for those data sets that contained evidence of an epidemic change.

2 Change-Point Detection Procedures for a Single Functional Time Series

2.1 Change-Point Detection Procedures

In this section we detail change-point detection procedures for a mean change in functional observations $X_i(t), t \in \mathcal{Z}$, $i = 1, \ldots, n$, where $\mathcal{Z}$ is some compact set. This setting for independent (functional) observations with at most one change-point (AMOC) was investigated by Berkes *et al.* [4] as well as Aue *et al.* [2] and for specific weak dependent processes by Hörmann and Kokoszka [19]. We will also allow for dependency (in time) of the functional observations (using meta-assumptions in order to allow for a very general class of dependency) and focus on the model with an epidemic change, where after a certain time the mean changes back. For this model some theoretical results relating to the detection and estimation of changes are given in Aston and Kirch [1].

The epidemic model is given by

$$X_i(t) = Y_i(t) + \mu(t) + \Delta(t) 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}}, \tag{2.1}$$

where the mean functions before the change $\boldsymbol{\mu} = \mu(\cdot)$, the change $\boldsymbol{\Delta} = \Delta(\cdot)$ as well as the functional time series $\{Y_i(\cdot) : 1 \leqslant i \leqslant n\}$ are elements of $L^2(\mathcal{Z})$, that are $(a.s.)$ continuous, $0 < \vartheta_1 \leqslant 1$ marks the beginning of the epidemic change, while $\vartheta_1 \leqslant \vartheta_2 \leqslant 1$ marks the end of the epidemic change. $\boldsymbol{\mu}$, $\boldsymbol{\Delta}$ as well as ϑ_1, ϑ_2 are unknown. Furthermore we assume that the time series $\{Y_i(\cdot) : i \geqslant 1\}$ is centered, stationary and ergodic with

$$\mathrm{E} \, \|Y_1(\cdot)\|^2 = \int \mathrm{E}(Y_1^2(t)) \, dt < \infty.$$

This compares to the AMOC-Model, which is given by

$$X_i(t) = Y_i(t) + \mu(t) + \Delta(t) 1_{\{\vartheta n < i \leqslant n\}}, \tag{2.2}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\Delta}$ and $\{Y_i(\cdot) : 1 \leqslant i \leqslant n\}$ are as above, $0 < \vartheta \leqslant 1$ describes the position of the change, $\mathrm{E}\, Y_i(t) = 0$. $\boldsymbol{\mu}$, $\boldsymbol{\Delta}$ as well as ϑ are unknown.

We are interested in testing the null hypothesis of no change in the mean

$$H_0 : \mathrm{E}\, X_i(\cdot) = \mu(\cdot), \quad i = 1, \ldots, n,$$

versus the epidemic change alternative

$$\begin{aligned} H_1 : &\mathrm{E}\, X_1(\cdot) = \mu(\cdot), \quad i = 1, \ldots, \lfloor \vartheta_1 n \rfloor, \lfloor \vartheta_2 n \rfloor + 1, \ldots, n, \quad \text{but} \\ &\mathrm{E}\, X_1(\cdot) = \mu(\cdot) + \Delta(\cdot) \neq \mu(\cdot), \quad i = \lfloor \vartheta_1 n \rfloor + 1, \ldots, \lfloor \vartheta_2 n \rfloor, \quad 0 < \vartheta_1 < \vartheta_2 < n. \end{aligned}$$

The null hypothesis corresponds to the cases where $\vartheta_1 = \vartheta_2 = 1$. All the arguments used in the rest of the paper hold analogously for the simpler case of AMOC changes, but as we expect epidemic changes in the application, we concentrate our focus on these.

It is well known how to test for mean changes in multivariate observations (cf. e.g. Horváth *et al.* [21]). However, in a functional setting or high-dimensional data this is computationally not feasible anymore. Here, the idea is to use a projection into a lower dimensional space and use standard change-point statistics for the projected data. Let this space be spanned by the orthonormal system $\{\widehat{v}_j(\cdot), j = 1, \ldots, d\}$, which usually

depends on the data and is obtained from some estimation procedure such as principle components.

Then, it holds for the projected data, under an epidemic change alternative ($i = 1, \dots, n, \quad l = 1, \dots, d$)

$$\widehat{\eta}_{i,l} := \langle X_i, \widehat{v}_l \rangle = \int X_i(t)\widehat{v}_l(t)\,dt = \langle Y_i, \widehat{v}_l \rangle + 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}} \langle \boldsymbol{\Delta}, \widehat{v}_l \rangle \tag{2.3}$$

In particular, $\widehat{\boldsymbol{\eta}}_i = (\widehat{\eta}_{i,1}, \dots, \widehat{\eta}_{i,d})^T$ is a d-dimensional time series exhibiting the same type of level shifts as the functional sequence $\{X_i(\cdot) : 1 \leqslant i \leqslant n\}$ if the change is not orthogonal to the subspace spanned by $\widehat{v}_1(\cdot), \dots, \widehat{v}_d(\cdot)$. Aston and Kirch [1] propose to use the following standard change-point statistics for an epidemic change on the projected data $\widehat{\boldsymbol{\eta}}_i = (\widehat{\eta}_{i,1}, \dots, \widehat{\eta}_{i,d})^T$:

$$\begin{aligned} T_n^{(A)} &= \frac{1}{n^3} \sum_{1 \leqslant k_1 < k_2 \leqslant n} \mathbf{S}_n\left(k_1/n, k_2/n\right)^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{S}_n\left(k_1/n, k_2/n\right), \\ T_n^{(B)} &= \max_{1 \leqslant k_1 < k_2 \leqslant n} \frac{1}{n} \mathbf{S}_n\left(k_1/n, k_2/n\right)^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{S}_n\left(k_1/n, k_2/n\right), \end{aligned} \tag{2.4}$$

where $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimator for the long-run covariance matrix $\boldsymbol{\Sigma} = \sum_{k \in \mathbb{Z}} \Gamma(k)$, $\Gamma(h) = \mathrm{E}\,\boldsymbol{\eta}_t \boldsymbol{\eta}_{t+h}^T$, $h \geqslant 0$, and $\Gamma(h) = \Gamma(-h)^T$ for $h < 0$ and

$$\mathbf{S}_n(x, y) = \sum_{nx < j \leqslant ny} \left(\widehat{\boldsymbol{\eta}}_j - \frac{1}{n} \sum_{i=1}^{n} \widehat{\boldsymbol{\eta}}_i \right).$$

For the small sample performance of the test the choice of estimator $\widehat{\boldsymbol{\Sigma}}$ is crucial, which is why we discuss some estimators in Section 3.1.

From (2.3) it is obvious that the choice of estimation procedure for the covariance structure and hence the basis function has a substantial influence under the alternative on the detectability of the change. In other words the behavior of this estimation procedure under alternatives is crucial for the power of the test. As a contrast the estimation procedure has only a very mild influence on the behavior under the null hypothesis. Precisely, the following mild conditions need to be fulfilled:

Assumption $\mathcal{ON}$. 1. Under the null hypothesis the estimated orthonormal system $\{\widehat{v}_k(\cdot), k = 1, \dots, d\}$ needs to stabilize in the following sense:

$$\int (\widehat{v}_k(t) - s_k v_k(t))^2\,dt = O_P(n^{-1}),$$

where $s_k = \operatorname{sgn}\left(\int v_k(t)\widehat{v}_k(t)\,dt\right)$ and $\{v_k(\cdot), k = 1, \dots, d\}$ is some orthonormal system. In particularly, $\{\widehat{v}_k(\cdot), k = 1, \dots, d\}$ is an estimator of $\{v_k(\cdot), k = 1, \dots, d\}$ up to the sign.

The above assumption is mainly an assumption on the estimation procedure and will be discussed in Section 2.3 where several possibilities to choose such a subspace are proposed. The following assumption is an assumption on the dependence structure of the original time series and is for example fulfilled for strong mixing time series as well as $L^p - m$-approximable time series (in the sense of Hörmann and Kokoszka [19]) due to the fact that $\{v_k(\cdot), k = 1, \dots, d\}$ is an orthonormal system (for more details we refer to Aston and Kirch [1]). Let $\eta_{i,l} = \int Y_i(t)v_l(t)\,dt$, $l = 1, \dots, d$, as well as $\boldsymbol{\eta}_i = (\eta_{i,1}, \dots, \eta_{i,d})^T$. In the following we assume:

- The time series $\{\boldsymbol{\eta}_i : i \in \mathbb{Z}\}$ is stationary and short-range dependent i.e.

$$\sum_{i \in \mathbb{Z}} |\operatorname{cov}(\eta_{0,l_1}, \eta_{i,l_2})| < \infty, \quad l_1, l_2 = 1, \dots, d.$$

- $\{\boldsymbol{\eta}_i\}$ fulfills the following functional limit theorem

$$\left\{ \frac{1}{\sqrt{n}} \sum_{1 \leqslant i \leqslant nx} \boldsymbol{\eta}_i \,:\, 0 \leqslant x \leqslant 1 \right\} \xrightarrow{D^d[0,1]} \{\boldsymbol{W}_d(x) : 0 \leqslant x \leqslant 1\},$$

where $\boldsymbol{W}_d$ is a d-dimensional Wiener process with positive-definite covariance matrix

$$\boldsymbol{\Sigma} = \sum_{k \in \mathbb{Z}} \Gamma(k), \qquad \Gamma(j) = \mathrm{E}\, \boldsymbol{\eta}_t \boldsymbol{\eta}_{t+h}^T \text{ for } h \geqslant 0, \text{ and } \Gamma(h) = \Gamma(-h)^T \text{ for } h < 0. \tag{2.5}$$

Under the above assumptions Aston and Kirch [1] prove the following asymptotics under H_0:

$$\begin{aligned} T_n^{(A)} &\xrightarrow{\mathcal{L}} \sum_{1 \leqslant l \leqslant d} \int \int_{0 \leqslant x < y \leqslant 1} (B_l(x) - B_l(y))^2 \, dx \, dy, \\ T_n^{(B)} &\xrightarrow{\mathcal{L}} \sup_{0 \leqslant x < y \leqslant 1} \sum_{1 \leqslant l \leqslant d} (B_l(x) - B_l(y))^2, \end{aligned} \tag{2.6}$$

where $B_l(\cdot)$, $l = 1, \ldots, d$, are independent standard Brownian bridges.

In order to obtain asymptotic power one for the above tests, the estimation procedure also needs to stabilize under alternatives, i.e.

Assumption $\mathcal{ON}$. 2. Let $\{w_k(\cdot), k = 1, \ldots, d\}$ be an orthonormal system, $\{\widehat{v}_k(\cdot), k = 1, \ldots, d\}$ the same estimators as before. Under the alternative we assume

$$\int (\widehat{v}_k(t) - s_k w_k(t))^2 \, dt = o_P(1),$$

where $s_k = \operatorname{sgn}\left(\int w_k(t) \widehat{v}_k(t) \, dt\right)$, i.e. the estimators converge to some contaminated ON-System. Note that w_k usually depends on the type of alternative.

As suggested by (2.3) we additionally need to assume that the change is not orthogonal to the contaminated ON-System, i.e. for some $k = 1, \ldots, d$ it holds

$$\int \Delta(t) w_k(t) \, dt \neq 0. \tag{2.7}$$

Then, Aston and Kirch [1] show that under the epidemic change alternative

$$T_n^{(A)} \xrightarrow{P} \infty, \qquad T_n^{(B)} \xrightarrow{P} \infty,$$

if $\widehat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\Sigma}_A$ for some symmetric positive-definite matrix $\boldsymbol{\Sigma}_A$. This shows, that the test has asymptotic power one. In particular, it becomes clear that it is mainly the estimation procedure to obtain an orthonormal basis for the projection that determines the power of the test.

As an estimator for the change-points, we propose

$$(\widehat{\vartheta}_1, \widehat{\vartheta}_2) = \arg\max \left(\mathbf{S}_n^T(x,y) \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{S}_n(x,y) : 0 \leqslant x < y \leqslant 1 \right), \tag{2.8}$$

where $\mathbf{S}_n(x,y)$ is as above and $(x_1, y_1) = \arg\max(Z(x,y) : 0 \leqslant x < y \leqslant 1)$ iff $x_1 = \min(0 \leqslant x < 1 : Z(x,y) = \max_{0 \leqslant s < t \leqslant 1} Z(s,t)$ for some $y)$ and $y_1 = \max(y > x_1 : Z(x_1, y) = \max_{0 \leqslant s < t \leqslant 1} Z(s,t))$. Aston and Kirch [1] prove that this estimator is consistent under the above assumptions and even get the following rate under slightly stronger assumptions:

$$(\widehat{\vartheta}_1, \widehat{\vartheta}_2) - (\vartheta_1, \vartheta_2) = O_P(n^{-1/2}). \tag{2.9}$$

2.2 Functional Magnetic Resonance Imaging: 1000 Connectome Resting State Data

We are now ready to take a first look at the data.

A resting state scan is one where an individual is asked to lie in the scanner for a period of time, usually with their eyes closed, and asked to think of nothing in particular while not falling asleep (see for example, Damoiseaux *et al.* [9]). Scans of this type are used to study the brain regions that are involved in underlying brain activity, also sometimes known as the default network, using various techniques many of which either explicitly or implicitly rely on stationarity of the time series (see Cole *et al.* [8] for an overview of the current methods of analysis and pitfalls associated with them). However, it is not known whether the areas just exhibit some stationary variation during the scan, or whether there are changes in activity during the scan that are more than could be expected just as a result of variability.

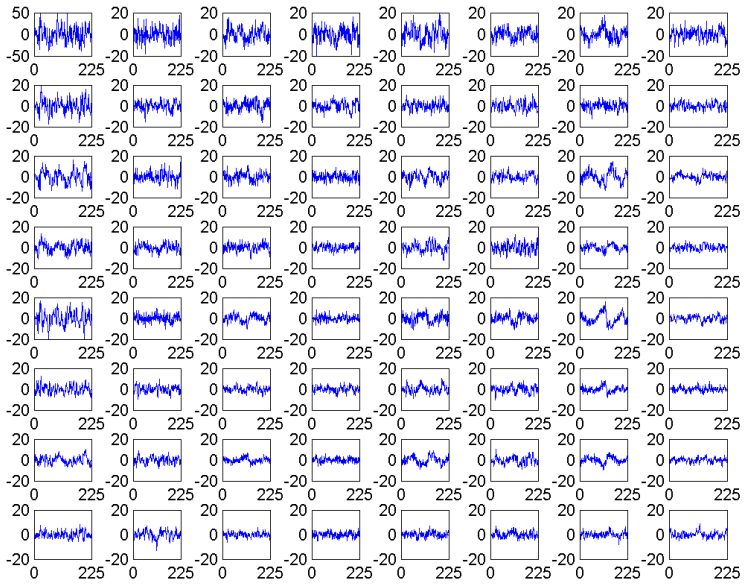
Activations in brain imaging are typically modelled as changes from baseline for a short period followed by a return to baseline (see for example Worsley *et al* [40]) showing that level shifts or change-point models describe well the kind of deviation from stationarity that can be expected. However, in resting state scans, it is not known when or even if any changes occur across time and thus change-point methods become more applicable than traditional experimental regression response type models. In addition, epidemic changes as the simplest model for multiple changes are a good first approximation to the deviation from stationarity. In fact, the premise of an epidemic change, where a return to baseline occurs, mimics the traditional activation / baseline response and will thus be of most interest here.

The data used here are publicly available from the 1000 Connectome project[1] [5]. This project consists of in excess of 1200 resting state data sets. However, a subset of this data will be used here so that confounding factors such as different scanner types and different locations of the subjects can be ignored. The data used was from a single site (Beijing China) and consists of 198 resting state scans, each consisting of 225 time points of a 3 dimensional image of size $64 \times 64 \times 33$ voxels with each temporal scan being taken 2 seconds apart (1 scan was discarded due to a different orientation of reconstruction, leaving 197 scans in the analysis below). Each scan had a polynomial trend of order 3 removed from each voxel time series prior to estimation to remove scanner drift (Worsley *et al.* [40]), in addition to being corrected for motion using the FSL software library (Jenkinson *et al.* [24]).

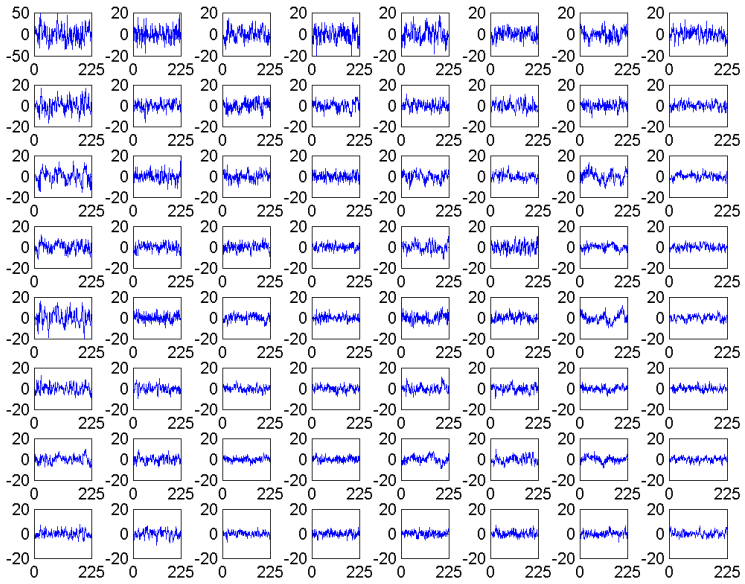
Since it is virtually impossible to examine the complete data set visually, in Figures 2.1 - 2.3, three data sets are shown after a separable dimension reduction to 64 dimensions was conducted (as described in Section 2.3.2 below). Recall that the original dimension is $64 \times 64 \times 33$ and therefore more than 2000 times as high.

The first subject in Figure 2.1 seems to exhibit strong deviations from stationarity – in fact the p-value associated with this time series is below 0.001 based on the bootstrap test as in Section 3.3 thus also survives the FDR correction. It should be stressed that the change detection is a global hypothesis test combined over all series considered. In this way, while taking more series will help increase the chance that the change is present in one series, it will come at the cost of the size of the change needed in finite samples for an omnibus test of this type. However, the subject shown in the figure did cause a rejection of the null hypothesis of no change both in the 64 and 125 subspace size omnibus tests, as well as surviving the multiple comparison correction due to the nearly 200 subjects considered. The components are shown both before and after

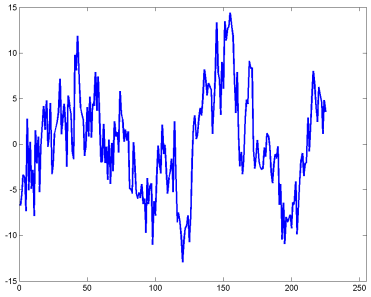
[1] These are publicly available data sets and can be accessed at `http://www.nitrc.org/projects/fcon_1000/`



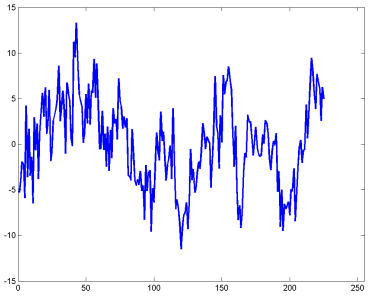
(a) 64 component time series



(b) 64 component time series - Epidemic changes removed

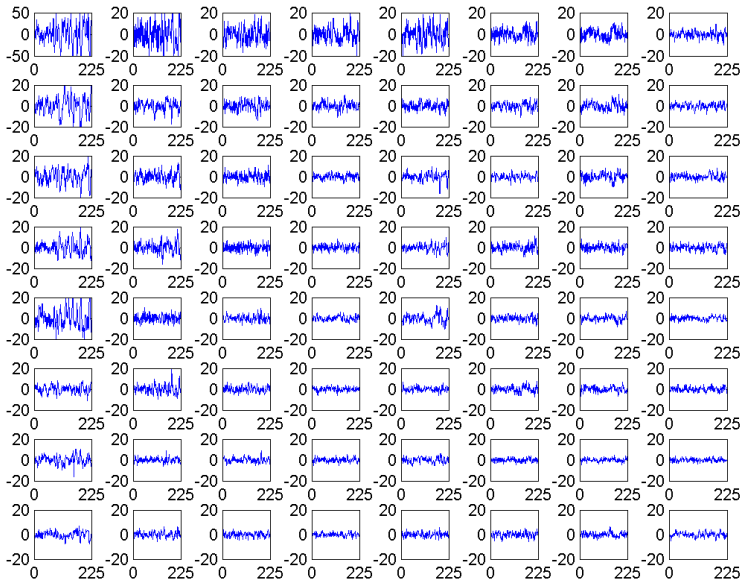


(c) Component 23 time series

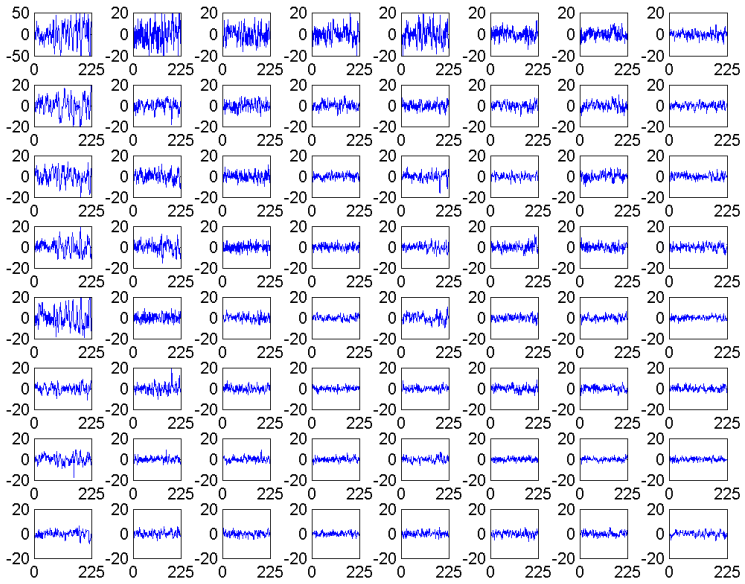


(d) Component 23: Epidemic change removed

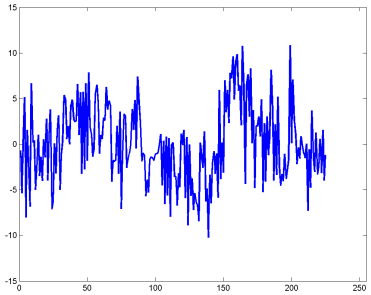
Figure 2.1: Subject 01018: Strong deviations from stationarity, $p < 0.001$



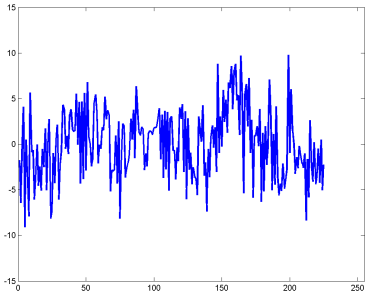
(a) 64 component time series



(b) 64 component time series - Epidemic changes removed



(c) Component 7 time series



(d) Component 7: Epidemic change removed

Figure 2.2: Subject 48501: Weak deviations from stationarity, $p < 0.05$, but not rejected when using FDR multiple comparisons correction

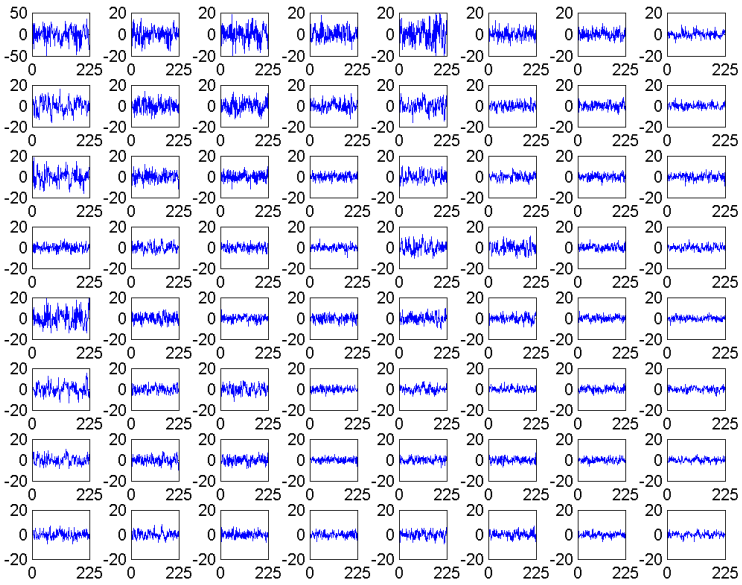


Figure 2.3: Subject 69518: No evidence of epidemic changes

the most likely change-points for each series are determined. In particular, the 23rd component is highlighted. This component can be seen to be a candidate series for a change to have occurred with the resulting change corrected series visually appearing much more stationary. The pictures indicate that an epidemic change is indeed a good first approximation but for this particular subject more deviation (maybe more change-points) seem to be present.

In Figure 2.2, a second subject is shown with a much smaller deviation from stationarity. In this case an epidemic change seems to be quite a good model for several components, but only a small part of the time series deviates from stationarity. For example, component 7 in Figure 2.2 shows a less pronounced but still plausible epidemic change compared with component 23 in Figure 2.1. The subject in Figure 2.2 also rejected the null hypothesis but only at about a 3% level, hence not surviving the FDR correction. Finally in Figure 2.3 a third subject is shown for which the components do not indicate level shifts and in fact the null hypothesis is not rejected for this subject.

2.3 Choice of Subspace for the Projection

The change-point tests described in Section 2.1 depend heavily upon the choice of estimation procedure for $\{\widehat{v}_k(\cdot), k = 1, \ldots, d\}$. In particular, it is required that the change is not orthogonal to the contaminated orthonormal system $\{w_k(\cdot), k = 1, \ldots, d\}$ which usually depends on the change and thus differs from the uncontaminated orthonormal system $\{v_k(\cdot), k = 1, \ldots, d\}$. In fact, this is a feature not a bug since a good choice of estimation procedure can have the nice property that any change detectable by the uncontaminated eigenspace is also detectable by the contaminated eigenspace and additionally the contaminated eigenspace $\{w_k\}$ differs from $\{v_k\}$ in such a way that the change is detectable using the contaminated eigenspace even though it would not have been detectable using the uncontaminated eigenspace (cf. Corollary 2.1).

2.3.1 Principal Components

Classical dimension reduction techniques are often based on the first d principle components, which choose a subspace explaining most of the variance. This procedure is also especially suited in the change-point situation but for completely different reasons. Heuristically speaking, standard variance estimators (such as the sample variance) increase in the presence of level shifts. Similarly, the variance estimate for linear combinations of components in the multivariate situation based on standard covariance matrix estimators will increase if a change is present in the linear combination. Thus, under the alternative, the principle components of the estimated covariance matrix will likely contain a change, so that assumption (2.7) is fulfilled. Corollary 2.1 formally proves this statement for the separable subspace selection if the change is separable. Note, however, that the corollary does not require that the true underlying covariance structure is separable for the statement still to be true. In the simpler situation of a general covariance structure and standard non-parametric covariance estimators an analogous assertion has been proven by Aston and Kirch [1]. Theorem 2.2 explains the situation for the separable estimation procedure for a general change. In this case, only a weaker result can be obtained.

To elaborate, consider the (spatial) covariance kernel of $Y_i(\cdot)$ given by

$$c(t,s) = \mathrm{E}(Y_i(t)Y_i(s)) \tag{2.10}$$

and define the covariance operator $C : \mathcal{L}^2(\mathcal{Z}) \to \mathcal{L}^2(\mathcal{Z})$ by $Cz = \int_{\mathcal{Z}} c(\cdot,s)z(s)\,ds$.

Let $\{\lambda_k\}$ be the non-negative decreasing sequence of eigenvalues of the covariance operator and $\{v_k(\cdot) : k \geqslant 1\}$ a given set of corresponding orthonormal eigenfunctions, i.e.

$$\int c(t,s)v_l(s)\,ds = \lambda_l v_l(t), \quad l = 1,2,\ldots, \quad t \in \mathcal{Z}. \tag{2.11}$$

More details can for example be found in either Bosq [6] or Horváth and Kokoszka [20]. The idea is now to choose the d eigenfunctions $v_l(\cdot)$, $l = 1,\ldots,d$, belonging to the largest d eigenvalues as a basis for the ON-System needed for the change-point procedure.

In practice, the covariance kernel $c(t,s)$ is usually not known, therefore it needs to be estimated. Even if $c(t,s)$ were known using estimators would be preferable due to the nice property that they can influence the contaminated system in such a way that the change becomes detectable (cf. Corollary 2.1).

Assumption $\mathcal{C}$. 1. Under H_0 the estimated covariance kernel $\widehat{c}_n(t,s)$ is a consistent estimator for the covariance kernel $c(t,s)$ of $\{Y_1(\cdot)\}$ with convergence rate $\sqrt{n}$, i.e.

$$\int\int (\widehat{c}_n(t,s) - c(t,s))^2\,dt\,ds = O_P(n^{-1}).$$

If one uses the principal components of $\widehat{c}_n$ as subspace for the change-point procedure, then $\mathcal{ON}.1$ holds if additionally the first $d+1$ eigenvalues of c fulfill $\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1} \geqslant 0$ (cf. Lemma 4.2 and 4.3 in Bosq [6]).

Assumption $\mathcal{C}$. 2. Under alternatives H_1 there exists a covariance kernel $k(t,s)$, such that

$$\int\int (\widehat{c}_n(t,s) - k(t,s))^2\,dt\,ds \xrightarrow{P} 0.$$

Usually the contaminated covariance kernel $k(t,s)$ as well as the contaminated eigenvalues γ_k will depend on the type and shape of the change. In fact this leads to the desirable property that a large enough change can influence k in such a way that it automatically is not orthogonal to the chosen subspace if the eigenfunctions belonging to the largest eigenvalues of $\widehat{c}_n$ are used (cf. Theorem 2.2 and in particular Corollary 2.1).

If one uses the principal components of $\widehat{c}_n$ as subspace for the change-point procedure, then $\mathcal{ON}.2$ holds if additionally the first $d+1$ eigenvalues of k fulfill $\gamma_1 > \gamma_2 > \ldots > \gamma_d > \gamma_{d+1} \geqslant 0$ (cf. Lemma 4.2 and 4.3 in Bosq [6]).

A natural estimator in a general non-parametric setting is the empirical version of the covariance function

$$\widehat{c}_n(t,s) = \frac{1}{n}\sum_{i=1}^{n}(X_i(t) - \bar{X}_n(t))(X_i(s) - \bar{X}_n(s)), \tag{2.12}$$

where $\bar{X}_n(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)$. In case of independent functional observations and for an AMOC change alternative Berkes *et al.* [4] proved $\mathcal{C}.1$ as well as $\mathcal{C}.2$ for this estimator. Their proof can be extended to the dependent AMOC situation (cf. Hörmann and Kokoszka [19]) as well as the dependent epidemic change situation (cf. Aston and Kirch [1]). For the latter the contaminated covariance kernel is given by

$$k(t,s) = c(t,s) + \theta(1-\theta)\Delta(t)\,\Delta(s), \qquad \theta = \vartheta_2 - \vartheta_1 > 0. \tag{2.13}$$

Usually one converts the continuous functional eigenanalysis problem to an approximately equivalent matrix eigenanalysis task. The simplest solution is a discretization of the observed function on a fine grid. Many data sets in applications are already obtained in this way as in the example of fMRI data used in this paper. For a discussion of this as well as more advanced options we refer to Ramsey and Silverman [33]. In such examples of very high-dimensional data, a principle component analysis based on the empirical covariance matrix is computationally infeasible due to the even higher-dimensionality of the covariance matrix. The following computational trick can be applied but also shows the limitations of the approach as the number of non-zero eigenvalues of the estimated covariance matrix is limited by the sample size, with the associated problems for small sample sizes.

Assume that after discretization the data is given by $X_i := (X_i(1), \ldots, X_i(M))^T$, $i = 1, \ldots, n$. The eigenanalysis problem corresponding to the estimated covariance kernel in (2.12) is to find the eigenvalues of the $M \times M$-matrix ZZ^T, where $Z = (X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$ is an $M \times n$-matrix. One can check that ZZ^T has $\operatorname{rank}(Z) \leqslant \min(M,n)$ non-zero eigenvalues which coincide with the $\operatorname{rank}(Z) \leqslant \min(M,n)$ non-zero eigenvalues of the $n \times n$- matrix Z^TZ. Furthermore the eigenvectors v_k of ZZ^T can be obtained from the eigenvectors u_k of Z^TZ by

$$v_k = \frac{Zu_k}{\|Zu_k\|}, \quad k = 1, \ldots, \operatorname{rank}(Z).$$

For more details we refer to Härdle and Simar [18, Ch 8.4]. Without presmoothing of the observed data it can easily happen that $M \gg n$ (as is the case in the fMRI data set we consider). In this case it is computationally much faster to calculate the eigenvectors of Z^TZ and then use the above transformation to obtain the eigenvectors of ZZ^T. This idea has been used for magnetic resonance imaging data in an i.i.d. setting in Zipunnikov *et al.* [41].

2.3.2 Principal Components Based on Separable Covariance Structures

The above discussion suggests that in many settings due to too a large a number of unknown parameters a loss of precision is unavoidable when the nonparametric

covariance estimator (2.12) is used. Therefore, in this section we assume a separable data structure which reduces the number of unknown parameters and can significantly improve computational speed as well as accuracy at least in situations where the data structure is correctly specified. If the covariance kernel is indeed separable, this approach leads to a correct estimation of the non-contaminated eigenspace under H_0 and to the estimation of a well-defined contaminated eigenspace under H_1. Even in the misspecified case, i.e. when the covariance kernel has a different structure, one estimates the basis functions of a well-defined subspace under both H_0 as well as H_1 but with a different interpretation (cf. Theorem 2.1).

For clarity of explanation, two dimensional data sets will be discussed here, although identical arguments apply for any finite number of dimensions. Indeed, the fMRI data set we consider is three dimensional so that a three-dimension version of the procedure below is used.

To this end consider the set $\mathcal{T} \times \mathcal{S}$, which is a product of two compact sets. Let $X_i(t,s), t \in \mathcal{T}, s \in \mathcal{S}$, $i = 1, \ldots, n$, and under H_0,

$$X_i(t,s) = Y_i(t,s) + \mu(t,s), \tag{2.14}$$

where the mean function $\mu(\cdot,\cdot)$ as well as the functional stationary time series $\{Y_i(\cdot,\cdot) : 1 \leqslant i \leqslant n\}$ are elements of $L^2(\mathcal{T} \times \mathcal{S})$, $\mathrm{E}\, Y_i(t,s) = 0$. Similarly, under alternatives

$$X_i(t,s) = Y_i(t,s) + \mu(t,s) + \Delta(t,s) 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}}.$$

The restricted covariance kernel of $Y_1(\cdot,\cdot)$ is assumed to fulfill

$$c((t_1,s_1),(t_2,s_2)) = c_1(t_1,t_2)c_2(s_1,s_2) \tag{2.15}$$

where $c_1(t_1,t_2)$ is an element of $L^2(\mathcal{T} \times \mathcal{T})$ and $c_2(s_1,s_2)$ an element of $L^2(\mathcal{S} \times \mathcal{S})$, with the full covariance function being an element of $L^2((\mathcal{T} \times \mathcal{S}) \times (\mathcal{T} \times \mathcal{S}))$. An important example of random data having such a separable structure is the following: Assume Y has mean 0 and covariance kernel $c_Y(t_1,t_2)$ independent of X, which has mean 0 and covariance kernel $c_X(s_1,s_2)$, then $Z(t,s) = Y(t)X(s)$ has covariance kernel $c_Y(t_1,t_2)c_X(s_1,s_2)$. In this example the data itself is separable from which the separability of the covariance as well as sample covariance kernel follows.

The factors c_1 and c_2 can only be obtained up to a multiplicative constant as

$$c((t_1,s_1),(t_2,s_2)) = (\alpha c_1(t_1,t_2)) \left(\frac{1}{\alpha} c_2(s_1,s_2)\right), \quad \alpha \neq 0,$$

but this does not cause a problem for the change-point procedures as will be seen below.

The eigenvalues λ_l resp. -functions v_l corresponding to c are the products of the eigenvalues $\lambda_{1,i}, \lambda_{2,j}$ resp. -functions $v_{1,i}, v_{2,j}$ of c_1 and c_2, since by (2.11)

$$\begin{aligned}
&\int_{\mathcal{T}} \int_{\mathcal{S}} c\left((t_1,s_1),(t_2,s_2)\right) v_{1,i}(t_2) v_{2,j}(s_2)\, dt_2\, ds_2 \\
&= \int_{\mathcal{T}} c_1(t_1,t_2) v_{1,i}(t_2)\, dt_2 \int_{\mathcal{S}} c(s_1,s_2) v_{2,j}(s_2)\, ds_2 \\
&= \lambda_{1,i}\, \lambda_{2,j}\, v_{1,i}(t_1)\, v_{2,j}(s_1).
\end{aligned} \tag{2.16}$$

We propose to use the subspace spanned by the products of the d_1 eigenfunctions belonging to the largest d_1 eigenvalues in the first dimension and the first d_2 eigenfunctions belonging to the largest d_2 eigenvalues in the second dimension. In a balanced situation it makes sense to choose $d_1 = d_2$ but sometimes there are fewer observations in one direction after discretization in which case $d_1 \neq d_2$ may be preferable. This

balanced choice of basis selection is preferable to choosing a basis of the eigenfunctions belonging the largest d joint eigenvalues as only then the eigenfunction will be guaranteed to include a large enough separable change (cf. Remark 2.1).

As in the non-parametric case one uses a discretized version of the covariance matrix for computations, so that this approach significantly reduces the computational complexity. For instance, if the observations consist of 100 data points in each direction, the covariance 'matrix' c is a $10\,000 \times 10\,000$ matrix while c_1 and c_2 are of dimension 100×100 each. The covariance matrix of a two-dimensional dataset Z can for example be obtained as the covariance matrix of $\tilde{Z} = \mathrm{vec}(Z)$, where vec is the operation that turns matrices into vectors by 'stacking' the columns. Under the above separability assumption, the covariance matrix of $\tilde{Z}$ corresponds to $c = c_1 \otimes c_2$, where $\otimes$ is the Kronecker product.

Separable covariance structures have obtained significant attention in the context of spatio-temporal statistics, where they have been used to separate the purely temporal covariance from the purely spatial covariance. While in our setup a temporal dependency is also present we use the separability approach only on the multidimensional spatial structure mainly for computational reasons to get a better and more stable approximation of the eigenfunctions in situations where the temporal sample is only moderately sized and the spatial structure is very high dimensional. In the context of spatio-temporal separability, several tests for a separable covariance structure have recently been developed which can also be applied in our situation (Fuentes [15] and Mitchell *et al.* [31]).

Furthermore, several approaches to estimate c_1 and c_2 from the data have also been discussed in the literature. Van Loan and Pitsianis [38] propose an algorithm which approximates a possibly non-separable covariance matrix by the closest (in the Frobenius norm) Kronecker product which has been shown to be useful in spatio-temporal covariance matrix approximation (Genton [16]). While this is a very appealing approach especially in view of misspecification, it is computationally not feasible in a high-dimensional context as it involves the calculation of singular vectors, which is computationally also very expensive. Dutilleul [13] proposes a MLE algorithm to estimate the factors, but again for high-dimensional data it is computationally too slow. However, their approach is related in the sense that they propose to start their algorithm with our estimator below. Extended and related algorithms have also been proposed for the estimation of separable covariance functions in a signal processing context (Werner *et al.* [39]) but are again designed for the use in small dimensional problems.

The estimated covariance kernel $\widehat{c}_n\left((t_1, s_1), (t_2, s_2)\right)$ as in (2.12) is used to estimate c_1 and c_2. Precisely consider

$$\widehat{c}_1(t_1, t_2) = \int_{\mathcal{S}} \widehat{c}_n((t_1, s), (t_2, s))ds \tag{2.17}$$

and

$$\widehat{c}_2(s_1, s_2) = \int_{\mathcal{T}} \widehat{c}_n((t, s_1), (t, s_2))dt. \tag{2.18}$$

In case of separability of c it holds

$$\widehat{c}_j(t_j, s_j) \xrightarrow{P} \frac{\operatorname{tr} c}{\operatorname{tr} c_j}\, c_j(t_j, s_j), \quad j = 1, 2,$$

where $\operatorname{tr} c(x, y) = \int c(x, x)\, dx$ and $\operatorname{tr} c = \sum_{i \geqslant 1} \lambda_i > 0$, if $c \neq 0$, where λ_i are the eigenvalues of the covariance operator $Cv = \int_{\mathcal{T}\times\mathcal{S}} c(\cdot, y)v(y)\, dy$ (cf. Theorem 4.1 in Gohberg *et al.* [17]) and analogously $\operatorname{tr} c_j > 0$. For the purpose of estimating the d

largest principal components this additional constant does not make a difference since the eigenfunctions are the same and the eigenvalues are only multiplied by a positive constant, thus not changing the order.

Next, we show that this leads to consistent eigenfunction estimation in the case, where the covariance kernel really is separable and at least to an orthonormal system in the misspecified case.

To this end assume that $\mathcal{C}.1$ respectively $\mathcal{C}.2$ hold, where c and k are kernels but not necessarily separable. The nonparametric covariance estimator (2.12) fulfills $\mathcal{C}.1$ as well as $\mathcal{C}.2$ for a large class of functional time series including mixing sequences or $L^p - m$-approximable time series (in the sense of Hörmann and Kokoszka [19]), for more details we refer to Aston and Kirch [1]. Let, under H_0,

$$\widetilde{c}_1(t_1,t_2) = \int_{\mathcal{S}} c((t_1,s),(t_2,s))ds, \quad \widetilde{c}_2(s_1,s_2) = \int_{\mathcal{T}} c((t,s_1),(t,s_2))dt,$$
$$\widetilde{c}\,((t_1,s_1),(t_2,s_2)) = \widetilde{c}_1(t_1,t_2)\,\widetilde{c}_2(s_1,s_2). \tag{2.19}$$

If the covariance kernel c is indeed separable i.e. fulfills (2.15) then $\widetilde{c}_j = \frac{\mathrm{tr}c}{\mathrm{tr}c_j}\,c_j$, $j = 1,2$ and $\widetilde{c} = \mathrm{tr}c\,c$. If the covariance kernel is not separable Theorem 2.1 shows that a subspace of the eigenspace of $\widetilde{c}$ is used for the change-point procedures.

Define $\widetilde{k}_1, \widetilde{k}_2, \widetilde{k}$ based on the contaminated covariance kernel $k((t_1,s_1),(t_2,s_2))$ analogously. Theorem 2.1 shows that a subspace of the eigenspace of $\widetilde{k}$ is used for the change-point procedure under alternatives. Thus all changes that are not orthogonal to this (contaminated) subspace are detectable (cf. (2.7) and following lines). However, $\widetilde{k}$ is not a multiplicative of k (as e.g. in (2.13)) as k is in general not separable not even if the covariance kernel c is separable. In fact, it will generally have a rather complicated shape making it difficult to derive relations between the eigenfunctions of $\widetilde{c}$ and those of $\widetilde{k}$. In the general situation Theorem 2.2 shows that most changes detectable by the corresponding uncontaminated basis remain detectable and in addition most large enough changes become detectable. Furthermore this statement is true for all separable changes (cf. Corollary 2.1).

Let

$$\widehat{v}_{(r,l)}(t,s) = \widehat{v}_{1,r}(t)\widehat{v}_{2,l}(s), \quad r = 1,\ldots,d_1, l = 1,\ldots,d_2, \tag{2.20}$$

where $\widehat{v}_{i,r}$ is the eigenfunction of $\widehat{c}_i$ as in (2.17) resp. (2.18) belonging to the rth largest eigenvalue. The following theorem shows that the non-contaminated subspace of this estimator is given as the corresponding subspace of eigenfunctions of $\widetilde{c}$ while the contaminated subspace is given as the corresponding subspace of eigenfunctions of $\widetilde{k}$.

Theorem 2.1. *a) Let $\mathcal{C}.1$ hold and in addition $\widetilde{\lambda}_{i,1} > \widetilde{\lambda}_{i,2} > \ldots > \widetilde{\lambda}_{i,d_i+1} \geqslant 0$, $i = 1,2$, where $\widetilde{\lambda}_{i,l}$ are the largest eigenvalues of $\widetilde{c}_i$. Let $v_{i,r}$ be the eigenfunctions of $\widetilde{c}_i$ belonging to the rth largest eigenvalue.*
Then $\widehat{v}_{(r,l)}(t,s)$ as in (2.20) and $v_{(r,l)}(t,s) = v_{1,r}(t)v_{2,l}(s)$, $r = 1,\ldots,d_1, l = 1,\ldots,d_2$, fulfill $\mathcal{ON}.1$.

b) Let $\mathcal{C}.2$ hold and in addition $\widetilde{\gamma}_{i,1} > \widetilde{\gamma}_{i,2} > \ldots > \widetilde{\gamma}_{i,d_i+1}$, $i = 1,2$, where $\widetilde{\gamma}_{i,l}$ are the largest eigenvalues of $\widetilde{k}_i$. Let $w_{i,r}$ be the eigenfunctions of $\widetilde{k}_i$ belonging to the rth largest eigenvalue.
Then $\widehat{v}_{(r,l)}(t,s)$ as in (2.20) and $w_{(r,l)}(t,s) = w_{1,r}(t)w_{2,l}(s)$, $r = 1,\ldots,d_1, l = 1,\ldots,d_2$, fulfill $\mathcal{ON}.2$.

Remark 2.1. Note that $v_{(r,l)}$ resp. $w_{(r,l)}$ are eigenfunctions belonging to the eigenvalue $\widetilde{\lambda}_{1,r}\widetilde{\lambda}_{2,l}$ of $\widetilde{c}$ resp. to the eigenvalue $\widetilde{\gamma}_{1,r}\widetilde{\gamma}_{2,l} \geqslant 0$ of $\widetilde{k}$ (cf. also (2.16)). This

'balanced' choice of basis is guaranteed to include at least the eigenfunctions belonging to the largest $d = \min(d_1, d_2)$ eigenvalues of the joint covariance kernel $\widetilde{c}$ (resp. $\widetilde{k}$). Furthermore, this is preferable to choosing just the eigenfunctions belonging to the largest d joint eigenvalues for the following reason: Corollary 2.1 b) shows that any large enough separable change has a tendency to switch the eigenfunctions in such a way that it becomes detectable. However, if the change is large in one of the components and small in the other one examples can be constructed where the change is not orthogonal to the d joint noncontaminated subspace but orthogonal to the contaminated subspace. The reason is that the contaminated eigenvalues in one component may become very large while only increasing slightly in the other component. As a result it is possible for all $\widehat{v}_{1,k} \cdot \widehat{v}_{2,l}$ corresponding to the joint largest eigenvalues that the first change $\mathbf{\Delta}_1$, e.g., is not orthogonal to $w_{1,k}$ but $\mathbf{\Delta}_2$ is orthogonal to $w_{2,l}$. But then $\Delta(t,s) = \Delta_1(t)\Delta_2(s)$ is orthogonal to the joint eigenspace belonging to the largest contaminated eigenvalues and hence not detectable. This cannot happen if one chooses a balanced basis of the products of the eigenfunctions belonging to the largest d_1 eigenvalues in the first dimension and largest d_2 eigenvalues in the second dimension (cf. Corollary 2.1 a)).

The following theorem characterizes the contaminated eigenfunction basis and its relation to the change. In part a) of the theorem a characterisation via the contaminated projection subspace is given for all changes that are detectable by the uncontaminated projection subspace, i.e. for which $\int_{\mathcal{T}} \int_{\mathcal{S}} \Delta(t,s) v_{1,r}(t) v_{2,l}(s)\, dt\, ds \neq 0$ for some r, l. Unlike in the situation where one uses a fully nonparametric covariance estimator to obtain the projection subspace (cf. Theorem 3.2 in Aston and Kirch [1]) we cannot conclude that the change is also detectable by the contaminated (separable) projection subspace, i.e. in general it does not hold that $\int_{\mathcal{T}} \int_{\mathcal{S}} \Delta(t,s) w_{1,r}(t) w_{2,l}(s)\, dt\, ds \neq 0$ for some r, l. Only the weaker statement given in the theorem can be derived. However, usually this will not cause a problem and the change will remain detectable. In the special case of a separable change, i.e. $\Delta(t,s) = \Delta_1(t)\Delta_2(s)$, detectability of a change by the uncontaminated projection subspace and the assertion obtained in the theorem are equivalent, showing that any separable change detectable by the uncontaminated projection subspace is also detectable by the contaminated subspace, which is the separable analogue of Theorem 3.2 a) from Aston and Kirch [1].

Part b) of the theorem gives a sufficient condition for which a large enough change is detectable using only a one-dimensional projection subspace. In case of a separable change this condition is fulfilled for any change, showing that any large enough separable change is detectable even when using only a one-dimensional subspace. This is the separable analogue of Theorem 3.2 b) in Aston and Kirch [1].

Note that it is not needed that c is separable.

Theorem 2.2. *a) Let $v_{j,r}$ be the rth largest eigenfunction belonging to $\widetilde{c}_j$ and $w_{j,r}$ be the rth largest eigenfunction belonging to $\widetilde{k}_j$ and let*

$$k\left((t_1,s_1),(t_2,s_2)\right) = c((t_1,s_1),(t_2,s_2)) + \theta(1-\theta)\Delta(t_1,s_1)\Delta(t_2,s_2)$$

as in (2.13)*. Then it holds:*

$$\begin{aligned} &\int_{\mathcal{T}} \int_{\mathcal{S}} \Delta(t,s) v_{1,r}(t) v_{2,l}(s)\, dt\, ds \neq 0, \quad \text{for some } 1 \leqslant r \leqslant d_1, 1 \leqslant l \leqslant d_2 \\ &\implies \int_{\mathcal{T}} \Delta(t,s) w_{1,r}(t)\, dt \not\equiv 0 \quad \text{for some } 1 \leqslant r \leqslant d_1 \\ &\qquad \text{and} \quad \int_{\mathcal{S}} \Delta(t,s) w_{2,r}(s)\, ds \not\equiv 0 \quad \text{for some } 1 \leqslant l \leqslant d_2. \end{aligned}$$

b) Let $\Delta_D(t,s) = D\Delta(t,s)$ *for some* $\Delta(t,s)$ *with* $\int_{\mathcal{T}}\int_{\mathcal{S}} \Delta^2(t,s)\,dt\,ds \neq 0$. *Let* $w_{j,k,D}$, $k = 1,\dots,d_j$, *be the normalized eigenfunctions belonging to the largest eigenvalues of the covariance kernel* $\widetilde{k}_{j,D}$ *obtained analogously to* (2.19) *with*

$$k_D\left((t_1,s_1),(t_2,s_2)\right) = c((t_1,s_1),(t_2,s_2)) + \theta(1-\theta)\Delta_D(t_1,s_1)\Delta_D(t_2,s_2).$$

Similarly, $x_{j,k}$, $j = 1,2$, *are the normalized eigenfunctions belonging to the largest eigenvalues* $\xi_{j,1} > \xi_{j,2} > \ldots > \xi_{j,d+1} \geqslant 0$ *of the kernel* $\int \Delta(t_1,s)\Delta(t_2,s)\,ds$ *resp.* $\int \Delta(t,s_1)\Delta(t,s_2)\,dt$.

Then, for $k = 1,\dots,d_j$, $j = 1,2$, *as* $D \to \infty$,

$$\|s_{j,k}w_{j,k,D}(\cdot) - x_{j,k}(\cdot)\| \to 0, \qquad s_{j,k} = sgn\left(\int w_{j,k,D}(z)x_{j,k}(z)\,dz\right).$$

In particular, there exists $D_0 > 0$ *such that*

$$\int \Delta_D(t,s)w_{1,k,D}(t)w_{2,l,D}(s)\,dt\,ds \neq 0, \quad 1 \leqslant k \leqslant d_1, 1 \leqslant l \leqslant d_2,$$

for all $|D| \geqslant D_0$, *if*

$$\int \Delta(t,s)x_{1,k}(t)x_{2,l}(s)\,dt\,ds \neq 0.$$

From the above discussion and theorem we obtain the following corollary for a separable change. Again note that we only need that the change is separable but not that the covariance structure is separable.

Corollary 2.1. *Assume that the change is separable, i.e.* $\Delta(t,s) = \Delta_1(t)\Delta_2(s) \not\equiv 0$.

a) Any change that is not orthogonal to the non-contaminated subspace is detectable:

$$\int_{\mathcal{T}}\int_{\mathcal{S}} \Delta_1(t)\Delta_2(s)v_{1,r}(t)v_{2,l}(s)\,dt\,ds \neq 0, \quad \text{for some } 1 \leqslant r \leqslant d_1, 1 \leqslant l \leqslant d_2$$
$$\Longrightarrow \int_{\mathcal{T}}\int_{\mathcal{S}} \Delta_1(t)\Delta_2(s)w_{1,r}(t)w_{2,l}(s)\,dt\,ds \neq 0, \quad \text{for some } 1 \leqslant r \leqslant d_1, 1 \leqslant l \leqslant d_2,$$

where the notation of Theorem 2.2 a) has been used.

b) With the notation of Theorem 2.2 b), there exists $D_0 > 0$ *such that*

$$\int \Delta_D(t,s)w_{1,1,D}(s)w_{2,1,D}(t)\,dt\,ds \neq 0$$

for all $|D| \geqslant D_0$. *This shows that any large enough change is detectable. In this case it even holds*

$$x_{j,1} = \pm\frac{\Delta_j(\cdot)}{\|\Delta_j(\cdot)\|}, \quad \text{i.e.} \quad \left\|\pm w_{j,1,D}(\cdot) - \frac{\Delta_j(\cdot)}{\|\Delta_j(\cdot)\|}\right\| \to 0$$

as $D \to \infty$.

It is clear that the choice of d_1 and d_2 plays an important role in terms of whether a change is detected or not. In principle component analysis frequently the number of components is chosen in such a way that 80% of the variability are explained. However, Corollary 2.1 b) suggests that a small number of components is often sufficient and may even increase the power. This leads to the approach described in the next section.

3 Practical Aspects of Small Sample Testing

3.1 Estimation of the Temporal Covariance Matrix

In the case where one deals with independent data and an estimation procedure that is – under the null hypothesis – capturing the true eigenfunctions of the covariance matrix, the long-run covariance matrix is diagonal. In this case only the variance of the scores need to be estimated which can easily be established using the sample variance.

On the other hand, if the data is dependent or one uses the separable estimation procedure on a non-separable covariance structure, estimation of the long-run covariance matrix $\mathbf{\Sigma}$ as in (2.5) is critical for the change-point procedure to yield reasonable results. However, this is a difficult task especially if the dimension of the projection subspace is large. For the separable estimation procedure this is typically the case, since it is a product of the dimension of the subspaces chosen in each component.

Most estimators for the long-run covariance matrix are based on

$$\widehat{\mathbf{\Sigma}} = \sum_{|h|\leqslant b_n} w_q(h/b_n)\widehat{\Gamma}(h),$$

for some appropriate weight function w_q and bandwidth b_n where $\widehat{\Gamma}(\cdot)$ is an estimator for the autocovariance matrix of the (uncontaminated) projected data vector. Hörmann and Kokoszka [19] prove consistency of this estimator for weakly dependent data. Politis [32] proposed to use different bandwidths for each entry of the matrix in addition to an automatic bandwidth selection procedure for the class of flat-top weight functions, where some additional modifications guarantee the estimate to be symmetric and positive definite. We follow his approach but adapt the estimator in such a way that it takes possible change-points into account thus improving the power of the test. For details in the univariate situation we refer to Hušková and Kirch [23].

Let

$$(\widehat{m}_{1,l}, \widehat{m}_{2,l}) = \arg\max_{k_1,k_2} \left(\left| \sum_{i=k_1}^{k_2} \widehat{\eta}_{i,l} - \frac{k_2-k_1}{n} \sum_{j=1}^{n} \widehat{\eta}_{i,l} \right| \right)$$

be the estimated change-points that are estimated separately in each component and let

$$\mathbf{e}(j) = (e_1(j), \ldots, e_d(j))^T, \tag{3.1}$$

$$\begin{aligned}
\text{where} \quad e_l(j) &= \widehat{\eta}_{j,l} - \bar{\widehat{\eta}}_{\widehat{m}_{1,l},\widehat{m}_{2,l}} 1_{\{\widehat{m}_{1,l}<j\leqslant\widehat{m}_{2,l}\}} - \bar{\widehat{\eta}}^{\circ}_{\widehat{m}_{1,l},\widehat{m}_{2,l}} 1_{\{j\leqslant\widehat{m}_{1,l},\widehat{m}_{2,l}<j\}}, \\
\bar{\widehat{\eta}}_{\widehat{m}_{1,l},\widehat{m}_{2,l}} &= \frac{1}{\widehat{m}_{2,l}-\widehat{m}_{1,l}} \sum_{j=\widehat{m}_{1,l}+1}^{\widehat{m}_{2,l}} \widehat{\eta}_{j,l}, \\
\bar{\widehat{\eta}}^{\circ}_{\widehat{m}_{1,l},\widehat{m}_{2,l}} &= \frac{1}{n-\widehat{m}_{2,l}+\widehat{m}_{1,l}} \sum_{1\leqslant j\leqslant\widehat{m}_{1,l},\widehat{m}_{2,l}<j\leqslant n} \widehat{\eta}_{j,l},
\end{aligned}$$

be the estimated uncontaminated data. Furthermore, we obtain an estimator of the uncontaminated autocovariance matrix as

$$\widehat{\Gamma}(h) = \frac{1}{n}\sum_{j=1}^{n-r} \widehat{\boldsymbol{e}}_j \widehat{\boldsymbol{e}}_{j+h}^T, \quad h \geqslant 0, \quad \widehat{\Gamma}(h) = \widehat{\Gamma}(-h), \quad h < 0.$$

We use the following flat-top kernel

$$w(t) = \begin{cases} 1, & |t| \leqslant 1/2, \\ 2(1-|t|), & 1/2 < |t| < 1, \\ 0, & |t| \geqslant 1, \end{cases}$$

and the bandwidth $B_{l,k} = B_{k,l} = 2\max(\widehat{b}_{l,k}, \widehat{b}_{k,l})$, where $\widehat{b}_{l,k}$ is the smallest positive integer such that

$$\left| \widehat{\Gamma}_{l,k}(\widehat{b}_{l,k}+j) / \sqrt{\widehat{\Gamma}_{l,l}(0)\widehat{\Gamma}_{k,k}(0)} \right| < 1.4\sqrt{\log_{10} n/n}, \qquad \text{for } j = 1, \ldots, 3.$$

Using the matrix of entries

$$\widehat{\mathbf{\Sigma}}^{(1)} = \left(\sum_{|h| \leqslant b_{k,l}} w_q(h/b_{k,l}) \widehat{\Gamma}_{k,l}(h) \right)_{k,l}$$

gives a symmetric estimator. However, while it is asymptotically positive definite, this is not necessarily true for small samples, it can even have negative eigenvalues. In fact, in the application the long-run covariance matrix had dimension 64×64 and estimation was conducted based on 225 data vectors only. The effect is that the estimation error can become rather large resulting in as much as thirty percent negative eigenvalues.

If one is only interested in a good positive definite estimator of the long-run covariance matrix but not its inverse as in our case, then the following approach yields reasonable results and was suggested by Politis [32] to overcome this deficiency. Consider the orthogonal diagonalization

$$\widehat{\mathbf{\Sigma}}^{(1)} = UDU^T$$

where U is an orthogonal matrix, $D = \operatorname{diag}(\delta_1, \ldots, \delta_d)$ with $\delta_1 \geqslant \delta_2 \geqslant \ldots \geqslant \delta_d$, and $D^+ = \operatorname{diag}(\delta_1^+, \ldots, \delta_d^+)$, $\delta_j^+ = \max(\delta_j, M_C/a_n)$, where $a_n \to \infty$ and

$$M_C = \text{Median}(m_1, \ldots, m_d),$$

where $m_1, \ldots, m_d$ are the eigenvalues of the estimated non-contaminated covariance matrix

$$\left(\frac{1}{n-1} \sum_{j=1}^{n} e_l(j) e_k(j) \right)_{k,l}.$$

This choice ensures that D^+ is scale invariant and asymptotically equal to D. A symmetric and positive definite estimator for $\mathbf{\Sigma}$ is thus given by

$$\widehat{\mathbf{\Sigma}}^{(2)} = UD^+U^T$$

However, for the change-point procedure in this paper an estimator for the inverse of $\mathbf{\Sigma}$ is needed, which can easily be obtained from the above decomposition as

$$\widehat{\mathbf{\Sigma}}^{(2)\,-1} = U(D^+)^{-1}U^T.$$

If the eigenvalues of the estimator are arbitrarily small, then the eigenvalues of the inverse become arbitrarily large, which in turns causes the change-point statistic to become very large. A cut-off point M_C/a_n as above solves the problem in principle,

but if it is chosen too small it will cause the change-point statistic to reject. For this reason, we used the following estimator in our data example for the inverse

$$\widehat{\mathbf{\Sigma}}^{(3)\,-1} = U\tilde{D}^{-1}U^T,$$
$$\text{where } \tilde{D}^{-1} = \operatorname{diag}\left(\widetilde{\delta_1^{-1}}, \ldots, \widetilde{\delta_d^{-1}}\right),$$
$$\widetilde{\delta_j^{-1}} = \begin{cases} \delta_j^{-1}, & \delta_j \geqslant M_C/a_n, \\ 0, & \text{else.} \end{cases}$$

This is a conservative estimator because the value of the change-point statistic using this estimator will be smaller than if δ_j had been set to any fixed small value. In the application a_n was set equal to $\log n$ and in all cases $M_c/a_n \approx 0.5$. Nevertheless even with this cut-off point conservative estimator the null hypothesis of stationarity was rejected for all subjects in our data example. We believe that this is due to the following fact:

If the dimension of the projection subspace is large in comparison to the length of the time series or if the time series deviates from stationarity in a different way than exhibiting an epidemic change, this estimator does not perform satisfactory. In this case it leads to a more stable and conservative change detection procedure if one only corrects for the long-run variance, setting all non-diagonal elements of the matrix equal to zero. Precisely, we use the following test statistics where $\widehat{\mathbf{\Sigma}}$ in $T_n^{(A)}$ resp. $T_n^{(B)}$ in (2.4) is replaced by $\widetilde{\mathbf{\Sigma}}$,

$$\widetilde{T}_n^{(A)} = \frac{1}{n^3} \sum_{1 \leqslant k_1 < k_2 \leqslant n} \mathbf{S}_n\left(k_1/n, k_2/n\right)^T \widetilde{\mathbf{\Sigma}}^{-1} \mathbf{S}_n\left(k_1/n, k_2/n\right),$$
$$\widetilde{T}_n^{(B)} = \max_{1 \leqslant k_1 < k_2 \leqslant n} \frac{1}{n} \mathbf{S}_n\left(k_1/n, k_2/n\right)^T \widetilde{\mathbf{\Sigma}}^{-1} \mathbf{S}_n\left(k_1/n, k_2/n\right), \tag{3.2}$$

where

$$\widetilde{\mathbf{\Sigma}}(i,i) = \widehat{\mathbf{\Sigma}}^{(2)}(i,i), \quad \widetilde{\mathbf{\Sigma}}(i,j) = 0 \text{ for } i \neq j. \tag{3.3}$$

is an estimator for the diagonal matrix of long-run variances:

$$V = (\gamma_i 1_{\{i=j\}})_{i,j=1,\ldots,d}, \qquad \gamma_i = \sum_{l \in \mathbb{Z}} \mathrm{E}\, \eta_{1,i}\eta_{1+l,i}.$$

Then, the limit distribution has still the same shape as in (2.6) but the Brownian bridges are no longer independent but rather exhibit the long-run correlation structure of the projected data. Furthermore, the results on the estimators (2.8) and (2.9) remain true. Since the limit distribution depends on unknowns, using asymptotic critical values is no longer feasible and the bootstrap introduced in the next section is essential.

3.2 Resampling Procedures for the Testing Problem

In practical applications it is often preferable to use resampling methods to obtain critical values rather than the asymptotic distribution. In small samples this can lead to improvements of size and power of the tests. In case of a non-pivotal limit distribution which one obtains for example when using the statistics $\widetilde{T}_n^{(A/B)}$ as in (3.2) asymptotic critical values differ from one time series to another so that resampling methods are the only way to obtain critical values. Permutation methods have the nice property that they are exact in an exchangeable situation, while bootstrap methods are only asymptotically exact but can have a better power if the null hypothesis is

correctly mimicked also under the alternative. For applications of the bootstrap to univariate change-point tests for dependent data we refer to Kirch [26] and Kirch and Politis [27].

In order to keep the procedure simple, we propose to use the following studentized circular block bootstrap (to allow for the time series error structure) taking a possible change-point separately in each component into account:
Let K be such that $n = KL$, $K, L \to \infty$, $K/L \to 0$.

(1) Let $e_l(j)$ be as in (3.1).

(2) Draw $U(1), \ldots, U(L)$ i.i.d., independent of $\{X(\cdot)\}$, such that $P(U(1) = i) = 1/n$, $i = 0, \ldots, n-1$.

(3) Let $e_l^*(Kj + k) := e_l(U(j) + k)$, $l = 1, \ldots, d$.

(4) Calculate

$$T_n^{(1)} := \frac{1}{n^3} \sum_{1 \leqslant k_1 < k_2 \leqslant n} \mathbf{S}_n^* \left(k_1/n, k_2/n\right)^T \widetilde{\mathbf{\Sigma}}^{*\,-1} \mathbf{S}_n^* \left(k_1/n, k_2/n\right),$$

$$\mathbf{S}_n^*(x, y) = \sum_{nx < j \leqslant ny} \left(e_j^* - \bar{e}_n^*\right), \qquad \bar{e}_n^* = \frac{1}{n} \sum_{i=1}^{n} e_i^*,$$

$$\widetilde{\mathbf{\Sigma}}^*(i,i) = \frac{1}{n} \sum_{l=1}^{L-1} \left(\sum_{k=1}^{K} (e_i^* \left(Kl + k\right) - \bar{e}_n^*) \right)^2, \quad \widetilde{\mathbf{\Sigma}}^*(i,j) = 0 \text{ for } i \neq j.$$

in case one wants to use statistic $\widetilde{T}_n^{(A)}$ and analogous versions for different statistics. Mark that the variance estimators used for the bootstrap are the block sample variances hence give the true variances of the conditional bootstrap distribution.

(5) Repeat steps (2)-(4) M times (e.g. $M = 1000$).

(6) $c^*(\alpha)$ is obtained as the upper α-quantile of $T_n^{(1)}, \ldots, T_n^{(M)}$.

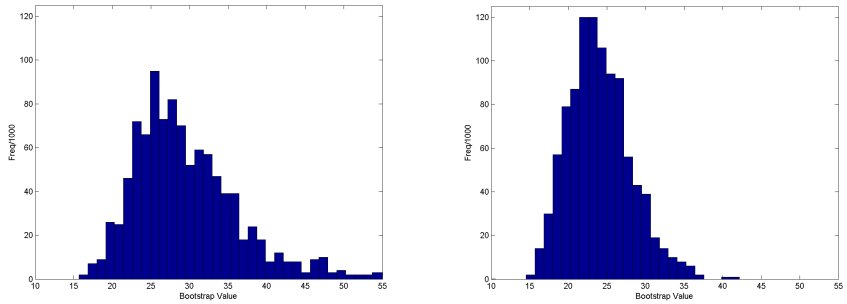
(7) Reject if $T_n > c^*(\alpha)$, where T_n is the statistic of interest, i.e. $\widetilde{T}_n^{(A)}$ in the above example, where one uses the estimator $\widetilde{\mathbf{\Sigma}}$ as given in (3.3).

A similar bootstrap has been applied by Kirch and Hušková [22, 23] in the univariate situation to obtain confidence intervals for the change-point. A proof for the validity of the univariate bootstrap (not taking possible changes into account) in the non-studentized case can be found in Kirch [25] under appropriate moment assumptions, extensions to the studentized case are immediate from (4.4) in Kirch and Hušková [23]. Extensions to the multivariate situation can be obtained along the same lines using Wolds Theorem. An additional problem in the situation in this paper is that $\widetilde{\eta}_{i,l}$ is not observed but needs to be estimated. Since only moment conditions of $\widetilde{\eta}_{i,l}$ are required for the proofs, extensions to $\widehat{\eta}_{i,l}$ are straightforward.

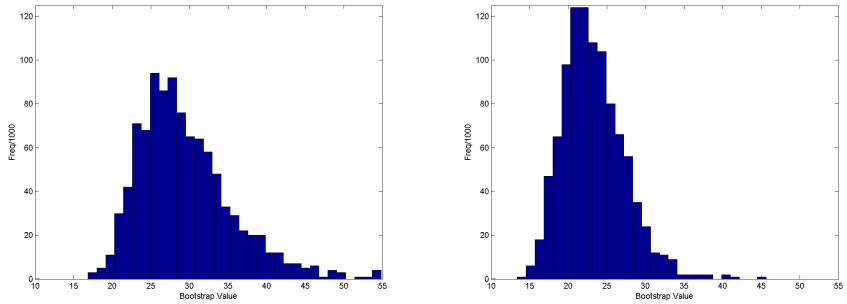
The choice of the block-length K is difficult – as a rule of thumb we propose to use $n^{1/3}$.

3.3 Testing for Epidemic Changes of the Connectome Data Set

In general, one of the biggest difficulties in analysing high dimensional multivariate data is the challenge of accurately estimating the covariance matrix. For our application, however, the biggest obstacle is to get an accurate estimate of the long-run



(a) Bootstrap distribution for two series with change detected



(b) Bootstrap distribution for two series with no change detected

Figure 3.1: Bootstrap distributions for 4 randomly chosen scans, two with changes detected, two with no changes detected, when using 125 components and the sum-statistic $\widetilde{T}_n^{(A)}$. The distributions vary due to the differing temporal correlation structures for different individuals.

temporal covariance of the projected data. As discussed in Section 3.1 obtaining a good estimate of the full long-run covariance matrix is highly problematic and all estimators discussed yield a poor performance when testing for changes in the Connectome data set. Therefore, we use the test statistics $\widetilde{T}_n^{(A/B)}$ as in (3.2) and the bootstrap critical values as described in Section 3.2 in the analysis of the data set.

Figure 3.1 shows four randomly chosen bootstrap distributions for statistic $\widetilde{T}_n^{(A)}$. As can be seen, the distribution support and shape depend on the correlation present within the data, but no visual evidence that there is any difference between distributions for scans which contain changes and those which do not.

After the preprocessing of the data described in Section 2.2, a functional principal component decomposition was used, based on the three orthogonal directions within the image acquisition. Eigen-decompositions of the empirical covariance functions were used to generate the full 3-dimensional functional basis. The eigenvalues associated with the decompositions did not decrease particularly fast. Indeed the first 1000 eigenvalues only explained approximately 5% of the variation. In many applications, this is unappealing as it means that the data cannot be sparsely represented. However, in change-point detection, a flat eigenstructure in the uncontaminated covariance can actually (and somewhat counter-intuitively) enhance detectability and is therefore actually an advantageous property. By Corollary 2.1 change-points, if present, will tend to be found in eigenfunctions with larger relative eigenvalues, and hence only a small number of components need to be checked especially when the components are flat. Thus, the number of components to examine was set to a small number, namely systems with 64 ($=4^3$) and 125 ($=5^3$) eigenfunctions were investigated, with each direction having either its top 4 or 5 eigenfunctions as part of the tensor product. This

Number of Components	Statistic used	Rejections (No Correction)	Rejections (FDR Correction)	FDR thresh
64	max ($\widetilde{T}_n^{(B)}$)	88	85	0.025
	sum ($\widetilde{T}_n^{(A)}$)	78	70	0.022
125	max ($\widetilde{T}_n^{(B)}$)	109	107	0.029
	sum ($\widetilde{T}_n^{(A)}$)	82	76	0.022

Table 3.1: Results of the 64 and 125 component analyses. 'No Correction' indicates all rejections at the 5% level were counted, while 'FDR Correction' indicates false discovery rate correction was used at a 5% level, with the corresponding threshold being given.

was a compromise between having a large number of components, which would reduce the finite sample detectability as well as computational speed (processing time for one scan with 1000 bootstrap samples for 125 components was approximately 6-7 hours on a desktop PC, while processing for the entire 197 scans took approximately 24 hours on a 40 node cluster), and having a sufficient number of components not to miss possible changes. Since the original data set was of dimension $64 \times 64 \times 33$ systems with 64 and 125 eigenfunctions correspond to an approximate dimension reduction by a factor of 2000 or 1000 respectively. Three examples (corresponding to strong, medium and no evidence for level shifts) of the projected data of dimension 64 are discussed in Section 2.2.

The use of separable functions for brain imaging is well known, either for smoothing (Worsley *et al.* [40]) or signal processing using techniques such as separable wavelets (Ruttimann *et al.* [36]).

The test statistics $\widetilde{T}_n^{(A/B)}$ in (3.2) were used to assess all 197 scans for a change-point. Bootstrap resampling as described in Section 3.2 was used to obtain critical values for each time series (M=1000). Multiple comparisons were corrected controlling the false discovery rate (FDR) by the procedure of Benjamini-Hochberg [3] for independent observations. In this case, unlike in usual brain imaging applications, the correction is done across subjects, not across space, as here space is a single functional observation, thus the observations (subjects) can be deemed independent.

The test results are summarized in Table 3.1. There was not a large difference whether 64 or 125 components were chosen, particularly for the sum statistic. Indeed, a small number of subjects became insignificant when 125 components instead of 64 components were used while others became significant. Therefore, the results look fairly stable regardless of the number of components chosen. If the sum statistic is used, approximately 40% of all subjects in the study were found to have some form of non-stationarity present which resulted in their being rejected as stationary against an epidemic alternative.

Bootstrap distributions from scans which contained detected changes looked very similar to those for scans without changes detected as shown in four typical examples in Figure 3.1. Figure 3.2 shows the distribution of the 5% bootstrap critical values from 197 scans, indicating that the critical values show some deviation between scans due to different underlying correlation structures hence different limit distributions, but do not differ between those with or without changes detected.

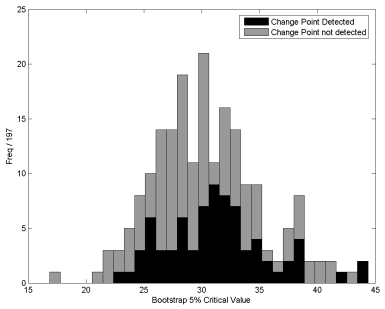


Figure 3.2: Distribution of bootstrap 5% critical values from 197 scans, where the stacking shows whether the critical value was from scan with detected or no detected change using 125 components and the sum-statistic $\widetilde{T}_n^{(A)}$.

4 Distribution of the Position and Duration of the Epidemic State

The discussion in the previous section has been dealing with situations, where one functional time series is observed and for this time series the question arises if and when a change has occurred.

In some situations, such as in psychological experiments or in stress testing due to the design of the experiment (cf. e.g. Lindequist *et al.* [30]), one can be reasonably sure that a certain change will occur. Usually in such situations more than one time series, namely one time series for each person involved in the experiment, is observed. Therefore it makes sense to include the change-point in the model and estimate the density of the change-point. For example one may be interested in knowing the distribution of the change-point in stress testing to get an idea about the change and duration distribution.

4.1 Density Estimation of the Change-Point for Hierarchical Time-Series

Let in case of AMOC

$$X_{i,j}(t) = Y_{i,j}(t) + \mu_j(t) + \Delta_j(t)1_{\{i>\vartheta_j n\}}, \quad 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m,$$

where the m observed functional time series $\{X_{1,j} : 1 \leqslant i \leqslant n\}, \ldots, \{X_{m,j} : 1 \leqslant i \leqslant n\}$ are independent, $\{\boldsymbol{\mu}_j : 1 \leqslant j \leqslant m\}$, $\{\boldsymbol{\Delta}_j : 1 \leqslant j \leqslant m\}$, and $\{\vartheta_j : 1 \leqslant j \leqslant m\}$ are no longer fixed deterministic but rather i.i.d. random variables independent of $\{Y_{i,j}(\cdot) : i \geqslant 1\}, j = 1, \ldots, m$, $P(0 < \vartheta_1 < 1) = 1$ and $P(\boldsymbol{\Delta}_1 \equiv 0) = 0$.

Furthermore we assume $n = n(m) \to \infty$ as $m \to \infty$.

Denoting $P^*(\cdot) = P(\cdot \,|\, \vartheta_j, \boldsymbol{\mu}_j, \boldsymbol{\Delta}_j, j = 1, \ldots, m)$ the consistency property $|\widehat{\vartheta} - \vartheta| = o_P(1)$ of AMOC estimators (cf. Theorem 2.3 in Aston and Kirch [1]) in the standard setting as outlined in Section 2.1 translates into:

$$|\vartheta_j - \widehat{\vartheta}_j| = o_{P^*}(1) \qquad a.s. \tag{4.1}$$

if the assumptions are $a.s.$ fulfilled, i.e. the mean changes are $a.s.$ non-orthogonal to the contaminated projection subspace and the basis is an orthonormal system almost surely.

Theorem 4.1. *If* (4.1) *holds and the distribution function* F_ϑ *of* ϑ *is continuous, then*

$$\widehat{F}_{\widehat{\vartheta},m}(x) := \frac{1}{m}\sum_{j=1}^{m} 1_{\{\widehat{\vartheta}_j \leqslant x\}}$$

is a consistent estimator for F_ϑ*, i.e.*

$$\sup_{x\in[0,1]} \left|\widehat{F}_{\widehat{\vartheta},m}(x) - F_\vartheta(x)\right| \to 0 \quad a.s.$$

The following theorem gives a corresponding results for kernel density estimators if a rate for the estimators of the change-point (analogously to (2.9)) is available.

Theorem 4.2. *Let* $h = h(m) \to 0, hm \to \infty$ *as* $m \to \infty$*. Assume*

$$h^{-1}|\vartheta_j - \widehat{\vartheta}_j| = o_{P^*}(1) \qquad a.s., \tag{4.2}$$

which follows for example from the analogue of (2.9) *if* $h^2 n \to \infty$*. Let* $K(\cdot)$ *be a bounded and Lipschitz continuous kernel (*$K(\cdot) \geqslant 0$*,* $\int K(x)\,dx = 1$*), then*

$$\int \mathrm{E}\left|\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x)\right|^2 dx \to 0,$$

where

$$\widehat{f}_{\widehat{\vartheta},m}(x) = \frac{1}{mh}\sum_{i=1}^{m} K\left(\frac{x-\widehat{\vartheta}_i}{h}\right)$$

and

$$\widehat{f}_m(x) = \frac{1}{mh}\sum_{i=1}^{m} K\left(\frac{x-\vartheta_i}{h}\right)$$

is the standard kernel estimator of the density f_ϑ *of* ϑ*.*

The theorem shows in particular that under standard assumptions on the kernel and the density it holds

$$\int \mathrm{E}\left|\widehat{f}_{\widehat{\vartheta},m}(x) - f_\vartheta(x)\right|^2 dx \to 0.$$

Remark 4.1. For the univariate problem one can show

$$P\left(\left|\widehat{\vartheta} - \vartheta\right| \geqslant c_n\right) \leqslant C(\min(\vartheta, 1-\vartheta))^{-2}\Delta^{-2}n^{-1}c_n^{-1},$$

where C does not depend on ϑ or μ, Δ, cf. e.g. Kokoszka and Leipus [28]. If additionally $\mathrm{E}[\Delta^{-2}\min(\vartheta_1, 1-\vartheta_1)^{-2}] < \infty$, then using the Markov-inequality and similar arguments as in the proofs of the above theorem one can conclude

$$\sup_x \left|f_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x)\right| \to 0 \qquad a.s.,$$

if e.g. $nh^3, mh^3 \to \infty$. This shows that in this situation under standard assumptions it holds $\sup_x \left|f_{\widehat{\vartheta}}(x) - f_\vartheta(x)\right| \to 0 \qquad a.s.$

If we are interested in estimators for an epidemic change things become slightly more complicated. The above results carry over immediately to $\widehat{\vartheta}_i = \widehat{\vartheta_{1i}}$ as an estimator for the first change-point as well as to $\widehat{\tau}_i = \widehat{\vartheta_{2i}} - \widehat{\vartheta_{1i}}$ as an estimator for the duration of the

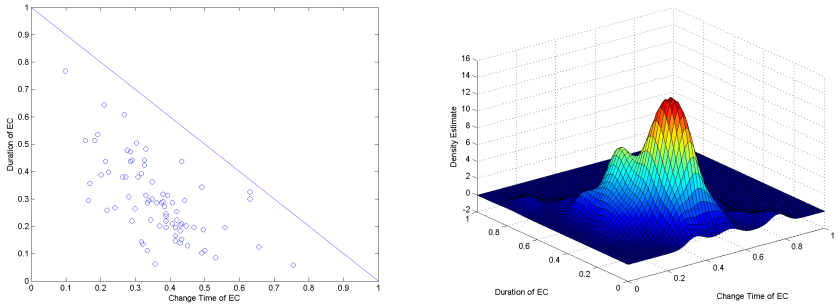


Figure 4.1: Estimators for 76 fMRI scans surviving FDR correction based on 125 components and the sum statistic $\widetilde{T}_n^{(A)}$.
Left: Joint estimates of position and duration of epidemic change.
Right: Kernel density estimate using a Gaussian kernel and bandwidths $h_x = 0.04, h_y = 0.05$.

epidemic change, so the marginal distributions can be estimated this way. Lindquist *et al.* [30] solve the problem by assuming that the first change-point ϑ_{1i} and the duration of the epidemic change τ_i are independent.

If one does not want to make this assumption, one can still formulate an analogous result to Theorem 4.2 using a two dimensional kernel $K(x, y)$, i.e. $\int K(x, y)dxdy = 1$, that is positive and bounded, and fulfills the following Lipschitz condition

$$|K(x_1, y_1) - K(x_2, y_2)| \leqslant C \max(|x_1 - x_2|, |y_1 - y_2|)$$

for some $C > 0$. Then, if $mh_1h_2 \to \infty$, $h_1, h_2 \to 0$, one gets an analogous result as in Theorem 4.2 for

$$\widehat{f}_{\widehat{\vartheta}_1,\widehat{\tau}_i,m}(x, y) = \frac{1}{mh_1h_2} \sum_{i=1}^{m} K\left(\frac{x - \widehat{\vartheta}_i}{h_1}, \frac{y - \widehat{\tau}_i}{h_2}\right),$$
$$\widehat{f}_m(x, y) = \frac{1}{mh_1h_2} \sum_{i=1}^{m} K\left(\frac{x - \vartheta_i}{h_1}, \frac{y - \tau_i}{h_2}\right).$$

The proof is analogous to the proof of Theorem 4.2.

4.2 Estimation for the Connectome Resting State Data

The results in the previous section can now be applied for the subjects that survived the FDR threshold as outlined in Section 3.3 and the joint distribution of position and duration of the epidemic change can be derived.

The left panel in Figure 4.1 shows the estimated change and durations for all those subjects where the null hypothesis of no change was rejected using FDR, while the right panel shows a kernel smoothed density estimate for the joint distribution of position and duration of the epidemic change, using the automatic bandwidth selection procedure of Botev *et al.* [7] (yielding bandwidths of $h_x = 0.04$, and $h_y = 0.05$). In this example change-points usually occur somewhere between 0.25 and 0.5, and last around 0.1-0.3 of the scanning period except for very early changes which often last longer. In fact, the density seems to be bimodal indicating two clusters dividing subjects into those for which a change occurs after a relatively short period in the scanner (maybe only now arriving in the stationary state) in addition to a relatively long duration

(possibly until the end of the scan), and those subjects for which after a short time in the epidemic state a return to baseline happens. Nevertheless it is important to note that for subjects with a relatively late change, a long duration cannot happen due to the limited time in the scanner, so this may also be an artefact of the statistical procedure meaning that subjects would remain in the epidemic state if they had been left in the scanner for a longer period of time.

The results of the study show that resting state scans in some cases do show evidence of deviation from stationarity which can be modelled by epidemic mean changes, at least as a first approximation, indicating that the overall activity is different at different times. This result has implications for studying correlations within the brain between regions of interest using multiple subjects, particularly if some subjects show non-stationary behaviour, while others do not.

5 Conclusions

In this paper, a methodology for the detection and estimation of change-points from multiple subjects has been outlined, and the associated statistical properties investigated. It has been shown that change-point analysis is a useful tool in situations where very high-dimensional data sets are collected across time, especially if the data has a natural spatial structure. One main result explains the impact of the choice of projection subspace estimation on the power of the tests. In particular, change-points will likely be found within the first few components when the eigenspectrum is relatively flat if one uses estimated principle components for the projection. The second main result shows that consistent estimators for the change-points exist and the associated distribution of change-point locations and durations can be found.

The aim of this paper was to find a general framework for the testing and estimation of change-points in resting state fMRI data, in such a way that details such as the estimation procedure for the projection subspace can be replaced with different statistical techniques while the underlying theoretical results remain valid. Examples include methodology based on fixed spatial basis choices such as wavelets, or computational methods such as those by Zipunnikov *et al.* [41] extended to non-i.i.d settings. For these variations, by careful choice of the estimators for the projection subspace, tests as well as estimators for the location and duration distributions can be obtained from the theoretic results given in this paper.

For future statistical analyses of resting state fMRI data, this study has three main implications:

- Firstly, routine testing for non-stationarities in resting-state scans is now possible, and relatively computationally inexpensive (compared to the time taken to do further analyses).

- Secondly, this study indicates that the examined subjects are fairly well split between those that have evidence of non-stationarities and those who do not, so that it would be of great interest to compare the connectivity relationships between these two groups. Many of the most standard connectivity measures are based on correlation analyses, which can be dramatically affected by the presence of non-stationarities. Hence, investigation of the phenomena found in this paper warrants further exploration.

- Thirdly, the distributions derived from the change-point estimators seem to indicate that the location and duration of the non-stationarities has considerable mass around half way through the scan. It would be of interest to investigate

further whether this is just the nature of the ability to rest within the scanner and is due to active thought processes interrupting the resting state network, or whether the resting state signal itself changes after a certain amount of time. This could be investigated by looking at the spatial distribution of the time series which exhibit changes, but requires further statistical development to rigorously allow the examination of individual spatial maps after the omnibus test for the presence of an epidemic change.

6 Proofs

Proof of Theorem 2.1. By the Cauchy-Schwarz inequality it holds

$$\begin{aligned} & n \int_{\mathcal{T}} \int_{\mathcal{T}} \left(\int_{\mathcal{S}} \widehat{c}_n((t_1,s),(t_2,s)) - c((t_1,s),(t_2,s))\, ds \right)^2 dt_1\, dt_2 \\ &= \int_{\mathcal{S}} 1\, ds\; n \int_{\mathcal{T}} \int_{\mathcal{T}} \int_{\mathcal{S}} \left(\widehat{c}_n((t_1,s),(t_2,s)) - c((t_1,s),(t_2,s))\right)^2 ds\, dt_1\, dt_2 \xrightarrow{P} 0, \end{aligned}$$

where the convergence follows from the assumptions of the theorem as well as the continuity of c and *a.s.* continuity of $\widehat{c}_n$. By Theorem 3.1 of Aston and Kirch [1] $\widehat{c}_1$, $\widetilde{c}_1$ fulfill Assumptions $\mathcal{C}.1$ and we can conclude that

$$\int_{\mathcal{T}} \left(\widehat{v}_{1,l}(t) - \mathrm{sgn}_{1,l} v_{1,l}(t)\right)^2 dt = O_P(n^{-1}),$$

where $\mathrm{sgn}_{1,l} = \pm 1$. Similarly one gets

$$\int_{\mathcal{S}} \left(\widehat{v}_{2,l}(s) - \mathrm{sgn}_{2,l} v_{2,l}(s)\right)^2 ds = O_P(n^{-1}).$$

Putting the two together yields an orthonormal system fulfilling $\mathcal{ON}.1$. The proof of b) is analogous. ■

Proof of Theorem 2.2. It holds

$$\widetilde{k}_1(t_1,t_2) = \widetilde{c}_1(t_1,t_2) + \theta(1-\theta) \int_{\mathcal{S}} \Delta(t_1,s)\, \Delta(t_2,s)\, ds,$$

Analogously

$$\widetilde{k}_2(s_1,s_2) = \widetilde{c}_2(s_1,s_2) + \theta(1-\theta) \int_{\mathcal{T}} \Delta(t,s_1)\, \Delta(t,s_2)\, dt.$$

We are now ready to prove a).

We prove the assertion by contradiction. Thus either

$$\int_{\mathcal{T}} \Delta(t,s) w_{1,r}(t)\, dt \equiv 0 \qquad \text{for all } r = 1,\dots,d_1,$$

or

$$\int_{\mathcal{S}} \Delta(t,s) w_{2,l}(s)\, ds \equiv 0 \qquad \text{for all } l = 1,\dots,d_2.$$

Analogous to the proof of Theorem 3.2 in Aston and Kirch [1] we obtain

$$\int_{\mathcal{T}} \Delta(t,s) v_{1,r}(t)\, dt \equiv 0 \qquad \text{for all } r = 1,\dots,d_1,$$

or

$$\int_{\mathcal{S}} \Delta(t,s) v_{2,l}(s)\, ds \equiv 0 \qquad \text{for all } l = 1, \ldots, d_2.$$

This in turn implies

$$\int_{\mathcal{T}} \int_{\mathcal{S}} \Delta(t,s) v_{1,r}(t) v_{2,l}(s)\, dt\, ds = 0, \qquad \text{for all } 1 \leqslant r \leqslant d_1, 1 \leqslant l \leqslant d_2.$$

For b) we show the result for $j = 1$, the assertion for $j = 2$ follows analogously. The eigenvectors of $\tilde{k}_1/D^2$ are the same as those of $\tilde{k}_1$, while the eigenvalues are also divided by D^2 hence remain in the same order. As $D \to \infty$ it holds

$$\begin{aligned} &\frac{\tilde{k}_1(t_1,t_2)}{D^2} = \frac{\tilde{c}_1(t_1,t_2)}{D^2} + \theta(1-\theta) \int_{\mathcal{S}} \Delta(t_1,s)\Delta(t_2,s)\, ds \\ &\to \theta(1-\theta) \int_{\mathcal{S}} \Delta(t_1,s)\Delta(t_2,s)\, ds. \end{aligned}$$

Hence, by Bosq [6], Lemmas 4.2 and 4.3, it holds

$$\pm\, w_{1,k,D} \to x_{1,k}, \quad k = 1, \ldots, d_1.$$

The second assertion of b) follows by the continuity of the scalar product in a Hilbert space. ■

Proof of Corollary 2.1. In case of separability it holds

$$\begin{aligned} &\int_{\mathcal{T}} \Delta(t,s) w_{1,r}(t)\, dt = \Delta_2(s) \int_{\mathcal{T}} \Delta_1(t) w_{1,r}(t)\, dt \not\equiv 0 \\ &\implies \int_{\mathcal{T}} \Delta_1(t) w_{1,r}(t)\, dt \neq 0. \end{aligned}$$

Analogously

$$\int_{\mathcal{S}} \Delta_2(s) w_{2,l}(s)\, ds \neq 0,$$

hence assertion a) follows from Theorem 2.2 a).

For b) note that in case of separability

$$\int_{\mathcal{S}} \Delta(t_1,s)\Delta(t_2,s)\, ds = \int_{\mathcal{S}} \Delta_2^2(s)\, ds\, \Delta_1(t_1)\Delta_1(t_2),$$

which has rank 1, i.e. only one non-zero eigenvalue. The corresponding eigenfunction is $\Delta_1(t_2)$. As a result $x_{j,1} = \pm\Delta_j/\|\Delta_j\|$, $j = 1, 2$. Since $\Delta \not\equiv 0$ it holds

$$\int_{\mathcal{T}} \int_{\mathcal{S}} \Delta(t,s)\Delta_1(t)\Delta_2(s)\, dt\, ds = \int_{\mathcal{T}} \int_{\mathcal{S}} \Delta_1(t)^2 \Delta_2(s)^2\, dt\, ds \neq 0,$$

so that b) follows from Theorem 2.2 b). ■

Proof of Theorem 4.1. Let $\widehat{F}_m(x) = \frac{1}{m} \sum_{i=1}^m 1_{\{\vartheta_i \leqslant x\}}$ be the (unobservable) empirical distribution function of F_ϑ. By the Glivenko-Cantelli lemma we know as $m \to \infty$

$$\sup_x |\widehat{F}_m(x) - F_\vartheta(x)| \to 0 \qquad a.s. \tag{6.1}$$

Hence it is sufficient to show that $\sup_x |\widehat{F}_{\widehat{\vartheta},m}(x) - \widehat{F}_m(x)| \to 0$ *a.s.* It holds

$$\left|\widehat{F}_{\widehat{\vartheta},m}(x) - \widehat{F}_m(x)\right| \leqslant \frac{1}{m}\sum_{i=1}^{n} 1_{\{A_i(x)\}},$$

where

$$A_i(x) = \begin{cases} \{\vartheta_i < x \leqslant \widehat{\vartheta}_i\}, & \vartheta_i \leqslant \widehat{\vartheta}_i, \\ \{\widehat{\vartheta}_i < x \leqslant \vartheta_i\}, & \vartheta_i > \widehat{\vartheta}_i. \end{cases}$$

For any $\epsilon > 0$ it holds

$$A_i(x) \subset \{|\vartheta_i - x| < \epsilon\} \cup \{|\widehat{\vartheta}_i - \vartheta_i| \geqslant \epsilon\},$$

hence

$$1_{\{A_i(x)\}} \leqslant 1_{\{|\vartheta_i - x| < \epsilon\}} + 1_{\{|\widehat{\vartheta}_i - \vartheta_i| \geqslant \epsilon\}}.$$

Consider $Y_m(i) := 1_{\{|\widehat{\vartheta}_i - \vartheta_i| \geqslant \epsilon\}} - P(|\widehat{\vartheta}_i - \vartheta_i| \geqslant \epsilon)$, a triangular array of rowwise i.i.d. centered random variables with $|Y_m(i)| \leqslant 2$ and

$$\mathrm{E}\left|\frac{1}{m}\sum_{i=1}^{m} Y_m(i)\right|^4 \leqslant C\frac{1}{m^2}$$

for some constant $C > 0$. An application of the Markov inequality as well as of the Borel-Cantelli Lemma thus shows $\frac{1}{m}\sum_{i=1}^{m} Y_m(i) \to 0$ *a.s.*, which implies in turn

$$\frac{1}{m}\sum_{i=1}^{m} 1_{\{|\widehat{\vartheta}_i - \vartheta_i| \geqslant \epsilon\}} \to 0 \qquad a.s.,$$

since by (4.1) and the dominated convergence theorem

$$P\left(|\widehat{\vartheta}_i - \vartheta_i| \geqslant \epsilon\right) = \mathrm{E}\, P^*\left(|\widehat{\vartheta}_i - \vartheta_i| \geqslant \epsilon\right) \to 0.$$

Noting that

$$1_{\{|\vartheta_i - x| < \epsilon\}} \leqslant 1_{\{\vartheta_i \leqslant x + \epsilon\}} - 1_{\{\vartheta_i \leqslant x - \epsilon\}},$$

an application of (6.1) shows that uniformly in ϵ

$$\begin{aligned} &\sup_x \left|\frac{1}{m}\sum_{i=1}^{m} 1_{\{|\vartheta_i - x| < \epsilon\}}\right| \\ &\leqslant \sup_x \left|\widehat{F}_m(x+\epsilon) - \widehat{F}_m(x-\epsilon)\right| \leqslant \sup_x |F_\vartheta(x+\epsilon) - F_\vartheta(x-\epsilon)| + o(1) \qquad a.s. \end{aligned}$$

which becomes arbitrarily small for $\epsilon \to 0$ by the uniform continuity of F_ϑ. The uniform continuity holds by the continuity and the fact that $F_\vartheta(x) = 0$ for $x \leqslant 0$ and $F_\vartheta(x) = 1$ for $x \geqslant 1$. Putting everything together yields the assertion. ∎

Proof of Theorem 4.2. Note that

$$\mathrm{E}\left|\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x)\right|^2 = \mathrm{var}\left(\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x)\right) + \left(\mathrm{E}\left(\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x)\right)\right)^2.$$

By the boundedness of K and since $\int \frac{1}{h} K\left(\frac{x-c}{h}\right)\, dx = \int K(x)\, dx = 1$ it holds for the first term

$$\begin{aligned}
&\int \operatorname{var}\left(\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x)\right)\, dx \\
&\leqslant \frac{1}{mh^2} \int \mathrm{E}\left(K\left(\frac{x-\widehat{\vartheta}_1}{h}\right) - K\left(\frac{x-\vartheta_1}{h}\right)\right)^2 dx \\
&\leqslant \frac{1}{mh}\left(\mathrm{E}\int \frac{1}{h} K\left(\frac{x-\widehat{\vartheta}_1}{h}\right) dx + \mathrm{E}\int \frac{1}{h} K\left(\frac{x-\vartheta_1}{h}\right) dx\right) = \frac{1}{mh} \to 0.
\end{aligned}$$

It holds

$$\int \left|\mathrm{E}\left(K\left(y\right) - K\left(y + \frac{\vartheta_1 - \widehat{\vartheta}_1}{h}\right)\right)\right| dy \leqslant 2,$$

as well as by the boundedness and Lipschitz-continuity of the kernel K

$$\begin{aligned}
&\left|\mathrm{E}\left(K\left(y\right) - K\left(y + \frac{\vartheta_1 - \widehat{\vartheta}_1}{h}\right)\right)\right| \\
&= O(1)\,\mathrm{E}\left(\min\left(1, \left|\frac{\vartheta_1 - \widehat{\vartheta}_1}{h}\right|\right)\right) \\
&= O(1)\,\mathrm{E}\left(1_{\{|\vartheta_1 - \widehat{\vartheta}_1| \geqslant \epsilon h\}}\right) + O(1)\,\mathrm{E}\left(\left|\frac{\widehat{\vartheta}_1 - \vartheta_1}{h}\right| 1_{\{|\vartheta_1 - \widehat{\vartheta}_1| < \epsilon h\}}\right) \\
&\leqslant \mathrm{E}\, P^*(|\vartheta_1 - \widehat{\vartheta}_1| \geqslant \epsilon h) + \epsilon = o(1) + \epsilon
\end{aligned}$$

by (4.2) and the dominated convergence theorem. We now conclude

$$\begin{aligned}
&\int \left(\mathrm{E}\left(\widehat{f}_m(x) - \widehat{f}_{\widehat{\vartheta},m}(x)\right)\right)^2 dx \\
&= \int \left[\frac{1}{h}\,\mathrm{E}\left(K\left(\frac{x-\vartheta_1}{h}\right) - K\left(\frac{x-\widehat{\vartheta}_1}{h}\right)\right)\right]^2 dx \\
&= \int \left[\mathrm{E}\left(K(y) - K\left(y + \frac{\vartheta_1 - \widehat{\vartheta}_1}{h}\right)\right)\right]^2 dy = o(1).
\end{aligned}$$

■

References

[1] Aston, J. A. D. and Kirch, C. Detecting and estimating epidemic changes in dependent functional data. *CRiSM Working Papers* 11-07, 2011.

[2] Aue, A., Gabrys, R., Horváth, L., and Kokoszka, P. Estimation of a change-point in the mean function of functional data. *J. Multivariate Anal.*, 100:2254–2269, 2009.

[3] Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57: 289–300, 1995.

[4] Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71:927–946, 2009.

[5] Biswal, Bharat B. *et al* Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107:4734–4739, 2010.

[6] Bosq, D. *Linear Processes in Function Spaces.* Springer, 2000.

[7] Botev, Z. I., Grotowski, J. F. and Kroese, D. P. Kernel density estimation via diffusion. *Ann. Statist.*, 38: 916–2957, 2010.

[8] Cole, D. M., Smith, S. M. and Beckmann, C. F. Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. Frontiers in System Neurosciencem, 4:8, 1–15, 2010.

[9] Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., and Beckmann, C. F. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103:13848–13853, 2006.

[10] Dehling, H. Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. verw. Geb.*, 63:393–432, 1983.

[11] Dehling, H. and Philipp, W. Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.*, 10:689–701, 1982.

[12] Dudley, R. M. and Philipp, W. Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Geb.*, 62:509–552, 1983.

[13] Dutilleul P. The MLE algorithm for the matrix normal distribution. *J. Statist. Comput. Simul.*, 64:105–123, 1999.

[14] Ferraty, F. and Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice.* Springer, New York, 2006.

[15] Fuentes, M. Testing for separability of spatial-temporal covariance functions. *J. Statist. Plann. Inference*, 136:447–466, 2004.

[16] Genton, M. G. Separable approximations of space-time covariance matrices. *Environmetrics*, 18:681–695, 2007.

[17] Gohberg, I., Goldberg, S., and Kaashoek, M. A. *Basic classes of linear operators.* Birkhäuser, Boston, 2003.

[18] Härdle, W. and Simar, L. *Applied Multivariate Statistical Analysis.* Springer, Germany, 2nd edition, 2007.

[19] Hörmann, S. and Kokoszka, P. Weakly dependent functional data. *Ann. Statist.*, 38:1845–1884, 2010.

[20] Horváth, L. and Kokoszka, P. *Inference for Functional Data with Applications.* 2010.

[21] Horváth, L., Kokoszka, P., and Steinebach, J. Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Multivariate Anal.*, 68:96–119, 1999.

[22] Hušková, M. and Kirch, C. Bootstrapping confidence intervals for the change-point of time series. *J. Time Ser. Anal.*, 29:947–972, 2008.

[23] Hušková, M. and Kirch, C. A note on studentized confidence intervals in change-point analysis. *Comput. Statist.*, 2009. To appear.

[24] Jenkinson M., Bannister, P. R., Brady, J. M. and Smith, S. M.. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17:825–841, 2002.

[25] Kirch, C. *Resampling Methods for the Change Analysis of Dependent Data.* PhD thesis, University of Cologne, Cologne, 2006. http://kups.ub.uni-koeln.de/volltexte/2006/1795/.

[26] Kirch, C. Block permutation principles for the change analysis of dependent data. *J. Statist. Plann. Inference*, 137:2453–2474, 2007.

[27] Kirch, C. and Politis, D. N. TFT-Bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *Ann. Statist.*, 2011 (To appear).

[28] Kokoszka, P., and Leipus, R. Change-point in the mean of dependent observations. *Statist. Probab. Lett.*, 40:385–393, 1998.

[29] Kuelbs, J., and Philipp, W. Almost sure invariance principles for partial sums of mixing b-valued random variables. *Ann. Probab.*, 8:1003–1036, 1980.

References

[30] Lindquist, M. A., Waugh, C., and Wager, T. D. Modeling state-related fMRI activity using change-point theory. *NeuroImage*, 35:1125–1141, 2007.

[31] Mitchell, M., Genton, M. G., and Gumpertz, M. Testing for separability of space-time covariances. *Environmetrics*, 16:819–831, 2005.

[32] Politis, D.N. Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. 2009. Preprint: Department of Economics, UCSD, Paper 2005-03R, http://repositories.cdlib.org/ucsdecon/2005-03R.

[33] Ramsay, J. O. and Silverman, B. W. *Functional Data Analysis*. Springer, Berlin, 2nd edition, 2005.

[34] Ranga Rao, R. Relation between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962.

[35] Robinson, L. F., Wager, T. D., and Lindquist, M. A. Change point estimation in multi-subject fMRI studies. *NeuroImage*, page in press, 2010.

[36] Ruttimann, U. E., Unser, M., Rawlings, R. R., Rio, D., Ramsey, N. F., Mattay, V. S., Hommer, D. W., Frank, J. A. and Weinberger, D. R. Statistical analysis of functional MRI data in the wavelet domain. *IEEE Trans. Med. Imaging* 17:142–54, 1998.

[37] Serfling, R.J. Convergence properties of S_n under moment restrictions. *Ann. Math. Statist.*, 41:1235–1248, 1970.

[38] Van Loan, C. F. and Pitsianis, N. Approximation with kronecker products. In M.S. Moonen and G.H. Golub, editors, *Linear Algebra for Large Scale and Real-Time Applications*, pages 293–314. Kluwer Publications, 1993.

[39] Werner, K., Jansson, M., and Stoica, P. On estimation of covariance martices with kronecker product structure. *IEEE Trans. Signal Processing*, 56:478–491, 2008.

[40] Worsley, K. J., Liao, C., Aston, J. A. D., Petre, V., Duncan, G. H. and Evans, A. C. A general statistical analysis for fMRI data. *NeuroImage* 15:1–15, 2002.

[41] Zipunnikov, V., Caffo, B., Crainiceanu, C., Yousem, D. M., Davatzikos, C. and Schwartz, B. S. Multilevel functional principal component analysis for high-dimensional data *Johns Hopkins University, Dept. of Biostatistics Working Papers.* Working Paper 219, 2010.